

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 22, 2005, 18:00:21 ; Search time 824 Seconds

(without alignments)
4580.514 Million cell updates/sec

Title: US-10-757-093-4

Perfect score: 3354
Sequence: 1 MKEFLGLSLSLAAPSISLGRP.....RKPKAAASHLRARMSIDKN 634

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 5544816 seqs, 297611598 residues

Total number of hits satisfying chosen parameters: 11089632

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.epool.p/US10757093/funat.18032005.164457.27782/app_query.fasta.1.775
-DB=Published Applications NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bases -START=1 -END=1 -MATRIX=blosum62
-TRANS=humand0.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=200 -THR MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEARSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10757093 -QCGN_1.1723 -funat.18032005.164457.27782
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBBLCK=100
-LONGLOG -DEV TIMEOUT=120 -WARP TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FEAPOP=6 -FEAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:
1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/prodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/prodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/prodata/2/pubpna/US09_PUBCOMB.seq.*
10: /cgn2_6/prodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/prodata/2/pubpna/US09D_PUBCOMB.seq.*
13: /cgn2_6/prodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq.*
15: /cgn2_6/prodata/2/pubpna/US10D_PUBCOMB.seq.*
16: /cgn2_6/prodata/2/pubpna/US10E_PUBCOMB.seq.*
17: /cgn2_6/prodata/2/pubpna/US10F_PUBCOMB.seq.*
18: /cgn2_6/prodata/2/pubpna/US10G_PUBCOMB.seq.*
19: /cgn2_6/prodata/2/pubpna/US10H_PUBCOMB.seq.*
20: /cgn2_6/prodata/2/pubpna/US11_NEW_PUB.seq.*
21: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq.*
22: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1717.5	51.2	3451	15	US-10-161-403-108
2	1717.5	51.2	3451	18	US-10-161-408-20
3	1717.5	51.2	14627	15	US-10-161-403-109
4	1717.5	51.2	14627	18	US-10-161-408-21
5	1715	51.1	8654	18	US-10-640-422-160
6	1714.5	51.1	8654	10	US-09-845-064-11
7	1714.5	51.1	8654	10	US-09-845-064-14
8	1714.5	51.1	9390	10	US-09-845-064-22
9	1714.5	51.1	9390	10	US-09-845-064-18
10	1714.5	51.1	9390	10	US-09-845-064-20
11	1714.5	51.1	9688	10	US-09-845-064-50
12	1714.5	51.1	15208	10	US-09-845-064-51
13	1714	51.1	10011	18	US-10-680-824A-19
14	1714	51.1	12224	16	US-10-332-566-83
15	1714	51.1	32798	16	US-10-424-638-1
16	1712	51.0	9285	10	US-09-845-064-52
17	1712	51.0	15077	10	US-09-845-064-57
18	1711	51.0	1809	18	US-10-432-777-16
19	1711	51.0	1812	15	US-10-161-403-105
20	1711	51.0	1812	17	US-10-363-493-24529
21	1711	51.0	1812	18	US-10-161-408-16
22	1711	51.0	11096	19	US-10-872-156-9
23	1709	51.0	1812	18	US-10-149-533A-13
24	1709	51.0	4652	10	US-09-893-525-36
25	1709	51.0	4652	18	US-10-763-380-36
26	1706	50.9	5390	10	US-09-893-525-41
27	1706	50.9	5390	18	US-10-763-380-41
28	1706	50.9	5418	10	US-09-893-525-38
29	1706	50.9	5418	18	US-10-763-380-38
30	1703	50.8	11978	10	US-09-792-568-8
31	1703	50.8	12438	10	US-09-792-568-9
32	1680.5	50.2	2755	17	US-10-239-907A-43
33	1680.5	50.1	4341	17	US-10-239-907A-47
34	1673.5	49.9	2141	17	US-10-663-241-27
35	1673.5	49.9	6975	17	US-10-431-252-17
36	1672.5	49.9	2001	17	US-10-239-907A-38
37	1672.5	49.9	2001	18	US-10-332-406A-21
38	1672.5	49.9	2001	18	US-10-800-161-29
39	1672.5	49.9	2730	17	US-10-239-907A-41
40	1672.5	49.9	4072	17	US-10-239-907A-49
41	1672.5	49.9	12817	18	US-10-332-406A-22
42	1672.5	49.9	13274	16	US-10-332-406A-20
43	1671.5	49.8	4129	18	US-10-332-656-28
44	1671	49.8	4280	15	US-10-027-880-2
45	1671	49.8	4309	15	US-10-027-880-4

ALIGNMENTS

RESULT 1
US-10-161-403-108
; Sequence 108, Application US/10161403
; Publication No. US20030119104A1
; GENERAL INFORMATION:
; APPLICANT: Perkins, Edward
; APPLICANT: Perez, Carl
; APPLICANT: Lindenbaum, Michael
; APPLICANT: Greene, Amy
; APPLICANT: Leung, Josephine
; APPLICANT: Fleming, Elena
; APPLICANT: Stewart, Sandra
; APPLICANT: Shellard, Joan
; TITLE OF INVENTION: CHROMOSOME-BASED PLATFORMS
; FILE REFERENCE: 24601-420
; CURRENT APPLICATION NUMBER: US/10/161,403
; CURRENT FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: 60/294,758
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: 60/366,891

```

; PRIOR FILING DATE: 2003-03-21
; NUMBER OF SEQ ID NOS: 129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 108
; LENGTH: 3451
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: HindIII Fragment containing the beta-glucuronidase
; OTHER INFORMATION: coding sequence, the rDNA intergenic spacer, and
; OTHER INFORMATION: the MastI sequence
US-10-161-403-108

Alignment Scores:
Pred. No.: 1,15e-169 Length: 3451
Score: 1717.50 Matches: 338
Percent Similarity: 68.84% Conservative: 95
Best Local Similarity: 53.74% Mismatches: 175
Query Match: 51.21% Indels: 21
DB: Gaps: 11

US-10-757-093-4 (1-634) x US-10-161-403-108 (1-3451)
QY 13 AlaAlaProSerLeuGlyThrProAlaAlaArgHisPheProArgAsnGlu----- 29
DB 1221 TCTGCACCGGATCTCCAGATC-----GAATTCCTCCGCGCGGGAATTCCTAG 1268
QY 30 --MetThrGlnHisGluGlnProLeuValLeuValArgProGlnHisThrSerArg 48
DB 1269 TGGATCCCGCGGATCGGTGCTTATG--TTACGTCCTGTAGAAACCCCAACCCG 1325
QY 49 GluLeuValAsnLeuAspGlyLeuThrPheAlaLeu--AlaSerGlyLeuAsnAsp 67
DB 1326 GAATCAAAAATTCAGACGGCTGTGGCATTCAGTGGATGCGAATACTGTGAATT 1385
QY 68 ThrAlaGlnProThrAlaProLeuProGlyLeuGluCysProValProAlaSer 87
DB 1386 GACACACGTTGGTGGAAACGGGTTACAAAGAACCGGCAATGCTGTGCCAGCAAG 1445
QY 88 TyrAsnAspIlePheIleSerArgGluIleHisAspHisValGlyTyrValTyrGln 107
DB 1446 TTTCACGATCAGTTCGCGCATGACGATATTCGTAATTATGTGGCAACGCTGTATCAG 1505
QY 108 ArgGluValIleValProGlySerGlyTyrPheGlnGluArgTyrLeuValArgAlaGlnSer 127
DB 1506 CGCGAAGCTTTTATACGAAAGGTTGGCGAGCCAGGCTATCGTCCGCTTCGATGCG 1565
QY 128 AlaThrHisHisGlyArgIleTyrValAsnAsnArgLeuValAlaGlnHisValGly 147
DB 1566 GTCACTCATTAACGCAAAAGTGGGTCAATATCAGGAAGTGAAGCATCAGGCGGCG 1625
QY 148 TyrThrProPheGluAlaAspValThrGlyLeuValAlaProGlyGlyCysPheArgLeu 167
DB 1626 TATATACGCAATTGAAAGCGATGTCACCGCGTATGTTATTCGCGGAAAGTGTACGTATC 1685
QY 168 ThrIleGlyValAsnAsnGluLeuThrHisGluThrIleProProGlyIleThrThr 187
DB 1686 ACAGTTTGTGTGAACACGAACTGAATGCGACAGCTATCCCGCGGAAAGTGTATACC 1745
QY 188 GlyAsnAlaThrGlyArgGlyIleGlnThrTyrGlnHisAspPheTyrAsnTyrAlaGly 207
DB 1746 ---GACAAACACGCAAGAAAGACAGTCTTATTCATGATTTCTTAACTACGCGGG 1802
QY 208 LeuAlaArgSerIleTyrPheTyrSerValProGlnGlnHisIleGlnAspIleThrVal 227
DB 1803 ATCCATGCGACGCGTAATGCTCTACACACCGCGAACCCTGGTGGACGATACACCGTG 1862
QY 228 ValThrAspVal-----AspGlyAspAsnGlyLeuIleAsnTyrGlyValAla 245
DB 1863 GTGACGATGTCGCGAAGCTGTACACACCGCTGTGACGCGAGTG---GTGGCC 1919
QY 246 AsnGlnThrThrGlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleValAla 265

```

```

DB 1920 AAT-----GGGATGTCAGCGTTGAACTGCGTGAATGCGGATCAACAGGTGTTCA 1970
QY 266 LysAlaSerGlyValaGlnGlyThrValThrIleProSerValIleLeuTyrGlnProGly 285
DB 1971 ACTGGAACAGGACCGACGGGACCTTGCAAGGTGAATCCGCACCTTGGCAACCGGCT 2030
QY 286 AlaAlaTyrLeuTyrGlnLeuGlnValAsnIleValGlySerSerGlyAspValAlaAsp 305
DB 2031 GAAGTTATCTCTAATGAACGTAGCTC-----ACAGCAAAAGCGACAGAGTGTAT 2084
QY 306 ThrTyrAsnLeuAlaThrGlyValaArgThrValIleAlaGlySerGlnPheLeuIle 325
DB 2085 ATCTACCGCGTGGCGGCGTCCGTCAGTGCAGTGAAGGGGGAACAGTTCCTGATC 2144
QY 326 AsnGlyIleProPheTyrPheThrGlyPheGlyHisGluAspThrAlaValaArgGly 345
DB 2145 AACCAACAAACGTTACTTATCTGCTTGGCGGTCAATGAAGATGCGGATTTGGCGGCG 2204
QY 346 LysGlyHisAspProAlaTyrMetValHisAspPheGlnLeuMetLysTyrIleGlyAla 365
DB 2205 AAGGATTCGATTAAGTGTGATGTGTCACGATCAACGATTAATGACTGATGGGCGC 2264
QY 366 AsnSerPheArgThrSerHisTyrProTyrAlaGluGluValaMetAspPheAlaAspArg 385
DB 2265 AACTCTTACCGTACCTCGCATTAACCTTAACGCTGAAGATGCTGACTGGCGACATGAA 2324
QY 386 AsnGlyIleValValIleAspGluThrProAlaValaGlyLeuAsnIleAlaLeu---Met 404
DB 2325 CATTGATGCTGTGATGATGATTAACCTGACAGTGTGCGCTTAACTCTCTTAAAGCAT 2384
QY 405 GlyValSerGlnSerGlyAlaPro---GlnThrPheThrProAspAlaIleAsnAspLys 423
DB 2385 GGTTCGAAGCGGCAACAGCGCAAGAACTGTACAGCAAGAGGACGTCACACGCGGAA 2444
QY 424 ThrGlnGluAlaHisLysGlnAlaIleArgGluLeuIleAlaArgAspLysHisAla 443
DB 2445 ACTCAGACCGCCACTTACAGCGATTAAGAGCTGATTAAGCGGTGACAAACCAACCCA 2504
QY 444 SerValIleMetTyrSerIleAlaAsnGluProAlaSerHisGluAspGlyAlaArgGlu 463
DB 2505 AGCGTGTATGATGAGATTTTCCCAACGAAACCGGATACCCGTCGCCAAGTGTCAACGGA 2564
QY 464 TyrPheGluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPhe 483
DB 2565 TATTTGCCCACTGGCGGAACGAACGCGTAACGATCCGACGCGTCCGATTCACCTGC 2624
QY 484 AlaSerValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPheAspLys 503
DB 2625 GTCAATGATATGTTCTGCACGCTCACACCGATACCATCAGCATCTCTTATGATGCTG 2684
QY 504 CysIleAsnArgTyrPheGlyTyrTyrSerGlnThrGlyAspLeuGluGluAla 523
DB 2685 TGCCTAACCGTTATTAACGTTGATGTCCAAAGCGGAGATTTGGAAACGCAAGAG 2744
QY 524 AlaLeuGluLysGluLeuHisGlyTyrGlnGlyLysPheHisArgProIleValIleThr 543
DB 2745 GTACTGAAAGAAAGAACTTGTGCGCTGGCAGAGAACTGCATCAGCGCATTTATCATCACC 2804
QY 544 GluTyrGlyAlaAspThrIleuAlaGlyLeuHisSerIleLeuGlyLeuProIlePheSer 563
DB 2805 GAATACGGGCGGTGATACGTTAGCGGGGTGCATCACTACACCGACATGTGAGTGA 2864
QY 564 GluPheGlnValaGlnMetLeuAspMetTyrHisArgValaPheAspArgIleGluSerMet 583
DB 2865 GAGTATCACTGTGATGCTGATGATATGATACCGCGTCTTTGATGCGTACGCGCGCTC 2924
QY 584 AlaGlyGlnHisValIlePheAsnPheAlaAspPheGlnThrAsnLeuGlyIleIleArgVal 603
DB 2925 GTGCGTGAACAGGTATGAAATTTCCGCAATTTGCAACCTCCGACAGGCAATATGTCGCGT 2984
QY 604 AspGlyAsnLysLysGlyValaPheThrArgAspArgLysProLysAlaIleHisSer 623
DB 2985 GCGGTGAACAAAGAGGATCTTACCCGCGACCGCAACCGAAAGTCGGCGGCTTTTCTG 3044

```

QY 624 LeuArgAlaArgThrPheSerIleAsp 632
 DB 3045 CTGCAAAAAGCTGGAGCTGGCATGAAC 3071

RESULT 2
 US-10-161-408-20
 / Sequence 20, Application US/10161408
 / Publication No. US20040214290A1
 / GENERAL INFORMATION:
 / APPLICANT: Perez, Carl
 / APPLICANT: Fabijanski, Steven
 / APPLICANT: Perkins, Edward
 / TITLE OF INVENTION: Plant Artificial Chromosomes
 / FILE REFERENCE: 24601-419
 / CURRENT APPLICATION NUMBER: US/10161,408
 / PRIOR FILING DATE: 2002-05-30
 / PRIOR APPLICATION NUMBER: US 60/294,687
 / PRIOR FILING DATE: 2001-05-30
 / PRIOR APPLICATION NUMBER: US 60/236,329
 / NUMBER OF SEQ ID NOS: 51
 / SOFTWARE: FastSeq for Windows Version 4.0
 / SEQ ID NO 20
 / LENGTH: 3451
 / TYPE: DNA
 / ORGANISM: Artificial Sequence
 / FEATURE:
 / OTHER INFORMATION: HindIII Fragment containing the beta-glucuronidase
 / OTHER INFORMATION: coding sequence, the rDNA intergenic spacer, and
 / OTHER INFORMATION: the MaetI sequence
 US-10-161-408-20

Alignment Scores:
 Pred. No.: 1,156-169 Length: 3451
 Score: 1717.50 Matches: 338
 Percent Similarity: 68.84% Conservative: 95
 Best Local Similarity: 53.74% Mismatches: 175
 Query Match: 51.21% Indels: 21
 Gaps: 11

US-10-757-093-4 (1-634) x US-10-161-408-20 (1-3451)

QY 13 AlaAlaProSerLeuGlyThrProAlaAlaArgHisPheProArgAsnGlu----- 29
 DB 1221 TCTGCACCGGATCTCGAGATC-----GAATTCCTCCGCGCGCGCAATTCAGTAG 1268

QY 30 ---MetThcGlnHisGluInProLeuIleValArgProGlnArgThrSerSerArg 48
 DB 1269 TGGATCCCGCGGATACGGTCCCTTATG---TTACGCTCTGTAGAAACCCCAACCGGT 1325

QY 49 GluLeuValAsnLeuAspGlyLeuTrpLysPheAlaLeu---AlaSerGlyLeuAsnAsp 67
 DB 1326 GAAATCAAAAACCTCGACGCGCTGTGGCATTTCAGTTCGATCGCAAACTGTGCAATT 1385

QY 68 ThrAlaGlnProTrpThrAlaProLeuProLysGlyLeuGluCysProValProAlaSer 87
 DB 1386 GAGCAGCGTGTGGGAAACCGCGCTTACAGAAACCGCGCAATTCGCTGTCCAGGCGAT 1445

QY 88 TyrAsnAspIlePheIleSerArgGluIleHisAspHisValGlyTrpValTyrTyrGln 107
 DB 1446 TTTAACGATACGATTCGCCCATGCAAGATATCGTAAATATATGGGCAACGCTGTGATCAG 1505

QY 108 ArgGluValIleValProLysGlyTyrPheGlnGluArgTyrLeuValArgAlaGluSer 127
 DB 1506 CCGGAAGCTTTTATACGAAAGGTGGGCGACGAGCGTANCGCTGCGTTTCGATGCG 1565

QY 128 AlaThrHisHisGlyArgGlyTyrValAsnAspArgLeuValAlaGluHisValGlyGly 147
 DB 1566 GTCACTCATTAATCGCAAAAGTGGGTCAATATATCAAGATGATGATGACATCAGGCGGC 1625

QY 148 TyrThrProheGluAlaAspValThrGluLeuValAlaProGlyGluLysPheArgLeu 167

DB 1626 TATACGCAATTTGAACCCATGTCACGCCGTATGTTATTCGCCGAAAGTGTACGATC 1685
 QY 168 ThrIleGlyValAsnAsnGluLeuThrHisGluThrIleProProGlyLysIleThr 187
 DB 1686 ACAGTTGTGTGAACAACGAACGAACTGAACCTGACACTATCCCGCGGAATGTGATTAAC 1745

QY 188 GlyAsnAlaThrGlyLysArgIleGlnThrTyrGlnHisAspPheTyrAsnTyrAlaGly 207
 DB 1746 ---GACGAAACCGCAAGAAAGAGCTTACTTCAATATTTCTTAATCAACGCCGG 1802

QY 208 LeuAlaArgSerIleTrpLeuTyrSerValProGlnGlnHisIleGlnAspIleThrVal 227
 DB 1803 ATCATTCGACGCGTAATAGCTTACACACGCGCAACCTGGGTGACGATATCACCGTG 1862

QY 228 ValThrAspVal-----AspGlyAspAsnGlyLeuIleAsnTyrGluValAla 245
 DB 1863 GTACACGATGTCGCGGACGCTGAACACGCGCTGTTCATCTGGCAGGTG---GTGGCC 1919

QY 246 AsnGlnThrThrGlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleValAla 265
 DB 1920 AAT-----GATATGTCACCGTTGAATCGTGATCGGATCAACAGGTGTTGCA 1970

QY 266 LysAlaSerGlyAlaGlnGlyThrValThrIleProSerValLysLeuTrpGlnProGly 285
 DB 1971 ACTGACAAAGGACACGCGGACCTTGCAAGTGATGATCGCACTCGCAACCGCGGT 2030

QY 286 AlaAlaTyrLeuTyrGlnLeuGlnValAsnIleValGlySerSerGlyAspValValAsp 305
 DB 2031 GAAGTTATCTTATACGAACTGATCGTC-----ACAGCCAAAGCCAGACAGTGTAT 2084

QY 306 ThrTyrAsnLeuAlaThrGlyValArgThrValLysValAlaGlySerGlnPheLeuIle 325
 DB 2085 ATCTACCGCTGCGCGCTCGCATCGGTCAGTGCAGTGAAGGCGCAACGTTCTGATC 2144

QY 326 AsnGlyLysProPheTyrPheThrGlyPheGlyLysHisGluAspThrAlaValArgGly 345
 DB 2145 AACCAAAACCGTTCTACTTACTTACGCTTGGCGCGTATGAATGCGGATTTGCCGCGC 2204

QY 346 LysGlyHisAspProAlaTyrMetValHisAspPheGlnLeuMetLysTrpIleGlyAla 365
 DB 2205 AAAGGATTCGATTAACGTCGATGATGTCACGATCAACGCAATTAATGCACTGATTTGGGCGC 2264

QY 366 AsnSerPheArgThrSerHisTyrProTyrAlaGluGluValMetAspPheAlaAspArg 385
 DB 2265 AACTCTACCGTACCTCGCATTAACCTTACGCTGAAGAGTGTGCACTGGGAGATGAA 2324

QY 386 AsnGlyIleValValIleAspGluThrProAlaValGlyLeuAsnIleAlaLeu---Met 404
 DB 2325 CATGCGATCGTGATGATGATGAACCTGACAGCTGTGCGCTTAACTCTTTAGCATT 2384

QY 405 GlyValSerGluSerGlyAlaPro---GlnThrPheThrProAspAlaIleAsnAspLys 423
 DB 2385 GGTTCGAAACCGCGCAACAAAGCTGAAGAACTGTATACGCAAGGACAGTCAACCGGGAA 2444

QY 424 ThrGlnGluAlaHisLysGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHisAla 443
 DB 2445 ACTCAGAGCGGACATTACAGGCGATTAAGACTGATTAACGCGGTGCAAAAACCAACCCA 2504

QY 444 SerValValMetTrpSerIleAlaAsnGluProAlaSerHisIleGluAspGlyAlaArgGlu 463
 DB 2505 AGCGTGATATGTCAGATATTCGCAAGAACCGGATACCGCTCGCAAGTGCACGGAA 2564

QY 464 TyrPheGluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPhe 483
 DB 2565 TATTTTCGCGCACTGCGCAAGCAACGCTGAATCGATCCGCGGTGCGATCACCTGC 2624

QY 484 AlaAsnValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPheAspValSer 503
 DB 2625 GTCAATGTAATGTTTCGCAAGCTCAACCGATACATCAAGCATCTTTGATATGCTG 2684

QY 504 CysIleAsnArgTyrPheGlyTyrTyrSerGlnThrGlyAspLeuGluAlaGluAla 523

```

Db      2685 TGCCGTAACCGTATTACGTTGATGTCCAAAGCGGCGATTGGAAACGCGACAGAG 2744
Qy      524  AAlaLeuGluYsGluLeuHhIGLYTTrGInGluYsPheHhISaGProIleValMetThr 543
Db      2745 GTACTGGAAAAAGAACTTCTGCGCTGGGAGAGAACTGCATCAAGCCGATATATATACC 2804
Qy      544  GltYrGlyAlaAspThrLeuAlaGlyLeuHhISerIleLeuGlyLeuProTrpSerGlu 563
Db      2805 GAATACGCGCTGGATACGTTAGCGGCTGCATCAATGATACACCGACATGTGAGTAA 2864
Qy      564  GltUheGInValGImetLeuAspMetYrHhISaGValPheAspArgIleGluSerMet 583
Db      2865 GAGATACGTTGATGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2924
Qy      584  AAlaGlyGluHhISaValTrpAspPheAlaAspPheGInThrAsnLeuGlyIleLeuArgVal 603
Db      2925 GTCCGTGAACAGATATGAAATTTTCCGATTTTCCGATTTTCCGATTTTCCGATTTTCCG 2984
Qy      604  AspGlyAsnIleYsGlyValPheThrArgAspArgIleYsProIleAlaAlaHhISer 623
Db      2985 GCGCGTAAACAGAGGCGATCTTCAACCGGACCGCAACGAAAGTCCGCGGCTTTCTG 3044
Qy      624  LeuArgAlaArgTrpThrSerIleAsp 632
Db      3045 CTGCAAAAACCGTGAATGCGCATGAAC 3071

```

RESULT 3
US-10-161-403-109
Sequence 109, Application US/10161403
Publication No. US20030119104A1
GENERAL INFORMATION:

APPLICANT: Perkins, Edward
APPLICANT: Perez, Carl
APPLICANT: Lindenbaum, Michael
APPLICANT: Greene, Amy
APPLICANT: Leung, Josephine
APPLICANT: Fleming, Elena
APPLICANT: Stewart, Sandra
APPLICANT: Sheilard, Joan
TITLE OF INVENTION: CHROMOSOME-BASED PLATFORMS
FILE REFERENCE: 24601-420
CURRENT APPLICATION NUMBER: US/10/161,403
CURRENT FILING DATE: 2002-05-30
PRIOR APPLICATION NUMBER: 60/254,758
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: 60/366,891
PRIOR FILING DATE: 2002-03-21
NUMBER OF SEQ ID NOS: 129
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 109

LENGTH: 14627
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: pag11a Plasmid
US-10-161-403-109

Alignment Scores:
Pred. No.: 9.01e-169 Length: 14627
Score: 1717.50 Matches: 338
Percent Similarity: 68.84% Conservative: 95
Best Local Similarity: 53.74% Mismatches: 175
Query Match: 51.21% Indels: 21
DB: 15 Gaps: 11

US-10-757-093-4 (1-634) x US-10-161-403-109 (1-14627)

```

Qy      13  AAlaAlaProSerLeuGlyThrProAlaAlaArgHhISaPheProArgAsnGlu----- 29
Db      12069 TCTGCAACCGGATTCAGATC-----GAATTCCTCCGCGCGCGCAATTCACTAG 12116
Qy      30  ---MetThrGlnHhISaGInProLeuIleYsValArgProGlnArgThrSerSerArg 48

```

```

Db      12117 TGGATCCCGCGGTAAGGTCAGTCCCTTATG---TTACGTCCTGTAGAAAACCCACCCGT 12173
Qy      49  GltLeuValAsnLeuAspGlyLeuTrpYsPheAlaLeu---AlaSerGlyLeuAsnAsp 67
Db      12174 GAATCAAAAACCTCAGACCGCTGCGGATTCAGCTGTGATCGCAAAACTGTGAAATT 12233
Qy      68  ThrAlaGlnProThrAlaProLeuProYsGlyLeuGluCyAspProAlaSer 87
Db      12234 GACACGCTGTGGTGGAAAGCGCTTACAAAGAAACCGGCAATTGCTGTGCCAGGCACT 12293
Qy      88  TyrAsnAspIlePheIleSerArgGluIleHhISaPheValGlyTrpValYrYrGln 107
Db      12294 TTAAAGCATAGTTCCCGCATGACATATTCGAATATTATGTGGCAACCTCGTATTCAG 12353
Qy      108  ArgGluValIleValProYsGlyTrpSerGlnGluArgYrLeuValArgAlaGluSer 127
Db      12354 CGGAAAGCTTTATACCGAAAGGTGGGAGGCGACCGCATGCTGTGTGCTGTGATGCG 12413
Qy      128  AlaThrHhISaGlyArgIleYrValAsnAspArgLeuValAlaGluHhISaValGlyGly 147
Db      12414 GTCACTCATTTACGCAAAAGTGGGTCAATTAATCAGAAAGTATGAGCATCAGGCGGC 12473
Qy      148  TyrThrProPheGlnAlaAspValThrGluLeuValAlaProGlyGluYsPheArgLeu 167
Db      12474 TATACCCCATTTGAAGCCGATGTCAAGCGTATGTTATGCGGGAAGATGATCGTATC 12533
Qy      168  ThrIleGlyValAsnAspGluLeuThrHhISaGluThrIleProProGlyYsAlaLeuThr 187
Db      12534 ACAGTTGTGTGAACAACAACTGAACCTGCGACATATCCCGCGGAATGTGATTCAC 12593
Qy      188  GltAsnAlaThrGlyYsArgIleGlnThrYrGlnHhISaPheThrYrAsnTrpAlaGly 207
Db      12594 ---GACGAAACCGGCAAGAAAGACGCTTACTTCATGATTTTAACTACGCCGG 12650
Qy      208  LeuAlaArgSerIleTrpLeuYrSerValProGlnGlnHhISaGlnAspIleThrVal 227
Db      12651 ATCCATCGACGAGGTATGCTTACACACCGCAACACCTGTGTGAGATATCAACGCTG 12710
Qy      228  ValThrAspVal-----AspGlyAspAsnGlyLeuIleAsnYrGluValGluValAla 245
Db      12711 GTGACGATGTGGCGGACGATACCAACGCGCTGTGTGACTGCGACAGTG---GTGCC 12767
Qy      246  AsnGlnThrThrGlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleValAla 265
Db      12768 AAT-----GGTATGTCAAGCGTGAACCTGCGTATGCGGATCAACAGGTGTTC 12818
Qy      266  YsAlaSerGlyAlaGlnGlyThrValThrIleProSerValYsLeuTrpGlnProGly 285
Db      12819 ACTGACAAAGGACCAAGCGGACCTTTCAGAGTGTGAATCCGACCTTGCACACCGGCT 12878
Qy      286  AAlaAlaYrLeuYrGlnLeuGlnValAsnIleValGlySerSerGlyAspValValAsp 305
Db      12879 GAAGTTATCTCTTATGAACGTGACCTC-----ACAGCAAAAGCCACAGAGTGTAT 12932
Qy      306  ThrYrAsnLeuAlaThrGlyValArgThrValYsValAlaGlySerGlnPheLeuIle 325
Db      12933 ATCTACCCGCTGCGGCGGATCGGATCGGATCGGATCGGATCGGATCGGATCGGATC 12992
Qy      326  AsnGlyYsProPheYrPheThrGlyPheGlyYsHhISaGluAspThrAlaValArgGly 345
Db      12993 AACCAAAACCGTTCTTACTTACCTGCTTGGCGGTCAAGATGCGGATTCGCGGC 13052
Qy      346  LysGlyHhISaPProAlaYrMetValHhISaPheGlnLeuMetYsTrpIleGlyAla 365
Db      13053 AAAGGATTCGATTAACGTGCTGATGTGACGATCAACGATTAATGAGATGATGGGCGC 13112
Qy      366  AsnSerPheArgThrSerHhISaYrProYrAlaGluGluValMetAspPheAlaAspArg 385
Db      13113 AACTCTAACCTTACTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTAC 13172
Qy      386  AsnGlyIleValIleAspGluThrProAlaValGlyLeuAsnIleAlaLeu---Met 404
Db      13173 CATGCACTGTGTGATGATGAACCTGCACTGTGCGCTTTTAACTCTTTAGGCAATT 13232

```



```
QY 405 GYVaisercglusergylalaPro---GlnThrPheThrProAspAlaIleAsnAspLeu 423
DB 13233 GGTTCGAAAGCCGGCAACAAGCCGAAAGAACTGTACAGCCAAAGGACGACGCGGAA 13292
QY 424 ThrGlnGlnAlaHisValysGlnAlaIleArgGluLeuIleAlaArgAspLeuAsnIleAla 443
DB 13293 ACTCAGACAGCGCAGCTTACAGGAGATTAAAGACTGATACGGGTGACAAAACCAACCA 13352
QY 444 SerValValMetTrpSerIleAlaAsnGluProAlaSerHisGluAspGlyAlaArgGlu 463
DB 13353 AGCGTGATGATGATGAGATATTCGCAACGAAACCGATCCCGTCCGAAAGTGACCGGAA 13412
QY 464 TyrPheGluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPhe 483
DB 13413 TATTTCGCCGACCTGGCGGAGCAAGCGGAACTGCATCCGACGCGCTCGATCACTGTC 13472
QY 484 AlaAsnValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPheAspValSer 503
DB 13473 GTCAATGTATATGTTCTGCGACGCTCACACCGATACATCAGCATCTTGTATGTGCTG 13532
QY 504 CysIleAsnArgTyrPheGlyTyrTyrSerGlnThrGlyAspLeuGluAlaGluAla 523
DB 13533 TGCCTAACCGTTATTAACGTTGATGTCGCAAGCGCGCATTTGGAAACGCGAGAGAG 13592
QY 524 AlaLeuGlnLysGluLeuHisGlyTyrGlnGlnLysPheHisArgProIleValMetThr 543
DB 13593 GATCTGAAAGAAACCTTCGCGCTCGCGAGAGAAATGATGATACCGCATTCATCACC 13652
QY 544 GluTyrGlyAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTyrPserGlu 563
DB 13653 GAATACGCGGTGATACGTTAGCGCGCTGCATCATGTACACCGACATGTGAGAGTGA 13712
QY 564 GluPheGlnValGlnMetLeuAspMetTyrHisArgValPheAspArgIleGluSerMet 583
DB 13713 GAGTATCAGTGTGATGCTGATGCTGATGATGATGATGATGATGATGATGATGATG 13772
QY 584 AlaGlnGlnHisValIleThrAsnPheAlaAspPheGlnThrAsnLeuGlyIleIleArgVal 603
DB 13773 GTCCGGGAAACAGGATGAGAAATTCGCGATTTTCGCGACGCGGATGATGCGGTT 13832
QY 604 AspGlyAsnLysLeuGlyValPheThrArgAspArgLysProLysValAlaAlaHisSer 623
DB 13833 GCGCGTAAACAAGAGGGGATCTTACCCGCGAGCCGAAACGGAAGTCGCGGCTTTCTG 13892
QY 624 LeuArgAlaArgTyrThrSerIleAsp 632
DB 13893 CTGCAAAAACGCTGACCTGCGCATGAAC 13919

RESULT 4
US-10-161-408-21
; Sequence 21, Application US/10161408
; Publication No. US20040214290A1
; GENERAL INFORMATION:
; APPLICANT: Perez, Carl
; APPLICANT: Pabjanek, Steven
; APPLICANT: Perkins, Edward
; TITLE OF INVENTION: Plant Artificial Chromosomes, Uses thereof, and Methods of Preparing
; FILE REFERENCE: 24601-419
; CURRENT APPLICATION NUMBER: US/10/161,408
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: US 60/294,687
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: US 60/296,329
; PRIOR FILING DATE: 2001-06-04
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PasteSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 14627
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
```

```
OTHER INFORMATION: pag1a Plasmid
US-10-161-408-21

Alignment Scores:
Pred. No.: 9,01e-169 Length: 14627
Score: 1717.50 Matches: 338
Percent Similarity: 68.84% Conservative: 95
Best Local Similarity: 53.74% Mismatches: 175
Query Match: 51.21% Indels: 21
DB: 18 Gaps: 11

US-10-757-093-4 (1-634) x US-10-161-408-21 (1-14627)

QY 13 AlaAlaProSerLeuGlyThrProAlaAlaArgHisPheProArgAsnGlu----- 29
DB 12069 TCTGACCCGAGATCTCAGATC-----GAATCCCGCGCGCGCGCAATTCTAG 12116
QY 30 ---MetThrGlnHisGluGlnProLeuIleLysValArgProGlnArgThrSerArg 48
DB 12117 TGGATCCCGCGGATACGCTGATCCCTTATG---TTAGTCTCTGTAGAAACCAACCGT 12173
QY 49 GluLeuValAsnLeuAspGlyLeuTyrLysPheAlaLeu---AlaSerGlyLeuAsnAsp 67
DB 12174 GAATCAAAAACCTCGACGCGCTGTGCGCATTCAGTCTGGATCGCAAACTGTGGAATT 12233
QY 68 ThrAlaGlnProThrAlaProLeuProLysGlyLeuGluCysProValProAlaSer 87
DB 12234 GACACCGTTGGTGGAAAGCGGTTACAAAGAAAGCCGCGCAATTGCTGTGCGAGCGAT 12293
QY 88 TyrAsnAspIlePheIleSerArgGluIleHisAspHisValGlyTyrValTyrTyrGln 107
DB 12294 TTTAACGATACGATCTCCCGCATGCGATATTCGTAATATGTGGGCAACGTGTGATACG 12353
QY 108 ArgGlnValIleValProLysGlyTyrPserGlnGluArgTyrLeuValArgAlaGluSer 127
DB 12354 CCGGAAGTCTTTATACCGAAAGGTTGGGCGACGACGATGCTGCTGCTTCGATGCG 12413
QY 128 AlaThrHisHisGlyArgIleTyrValAsnAsnGluValAlaGlnHisValGlyGly 147
DB 12414 GTCACTATTTACCGCAAGAGTGCGTCAATATACAGAAAGTATGAGCATCAGGCGGC 12473
QY 148 TyrThrProPheGlnAlaAspValThrGluLeuValAlaProGlyGluLysPheArgLeu 167
DB 12474 TATACCCGATTTGAACCCGATGTCACGCGGTATGTTATGCGGAAAGATGATACGATC 12533
QY 168 ThrIleGlyValAsnAsnGluLeuThrHisGluThrIleProProGlyLysIleThrThr 187
DB 12534 ACAGTTTGTTGTAACAACGAACTGAACCTGCGACACTATCCCGCGGAATGATGATTAAC 12593
QY 188 GlyAsnAlaThrGlyLysArgIleGlnThrTyrGlnHisAspPheTyrAsnTyrAlaGly 207
DB 12594 ---GACGAAACCGGCAAGAAAGAGTCTTACTTCATATATTTCTTAACTAGCGCGG 12650
QY 208 LeuAlaArgSerIleTyrLeuTyrSerValProGlnGlnHisIleGlnAspIleThrVal 227
DB 12651 ATCCATCGACGCTATATGCTTACACCGCGCAACACCTGGGTGACGATACACCGTG 12710
QY 228 ValThrAspVal-----AspGlyAspAsnGlyLeuIleAsnTyrGluValAla 245
DB 12711 GTGACCCAGTCCGCGAAGCTGTACACACGCTGTGTGACGTGCGAGG---GTGCGCC 12767
QY 246 AsnGlnThrThrGlyGlnIleGlnIleSerValIleAspGluAspArgValAlaIleValAla 265
DB 12768 AAT-----GATATGTCACGCTTGAACCTCGATGCGGATCAACAGTGTGTGCA 12818
QY 266 LysAlaSerGlyValaGlnGlyThrValThrIleProSerValLysLeuTyrGlnProGly 285
DB 12819 ACTGACACGACCAACGCGGACTTTGCAAGTGTGAATTCGCACTTGTGCAACCGGAT 12878
QY 286 AlaAlaTyrLeuTyrGlnLeuGlnValaAsnIleValaGlySerSerGlyAspValValAsp 305
DB 12879 GAAGTTATCTATGAACTGTACGTC-----ACAGCCAAAGCCAGACAGAGTGAT 12932
```

```

QY 306 ThrTyrAsnLeuAlaThrGlyValArgThrValLeuValAlaGlySerGlnPheLeuIle 325
Db 12933 ATCTACCGCGTGGCGCGCGATCCGCTCACTGCACTGAAAGCGCAAGCTTCTATC 12992
QY 326 AsnGlyLysProPheTyrPheThrGlyPheGlyLysIleGluPheThrAlaValArgGly 345
Db 12993 AACCAACAAACCGTCTACTTACTTACGCTTGGCGATGATGAGATGCGATTGGCGGCG 13052
QY 346 LysGlyIleAspProAlaTyrMetValHisAspPheGlnMetLeuIleTyrPheGlyAla 365
Db 13053 AAGAGATTCGATTAACGCTGATGATGTCACATCAACCATTAATGACTGATTTGGGGCC 13112
QY 366 AsnSerPheArgThrSerHisTyrProTyrAlaGluGluValMetAspPheAlaAspArg 385
Db 13113 AACTCCACCGTACTGCTGATACCTTACGCTGAAGATGCTCCGACCTGGCAGATGAA 13172
QY 386 AsnGlyIleValValIleAspGluThrProAlaValIleLeuAsnIleAlaLeu--Met 404
Db 13173 CATGGCATCGTGGATGATGAAACCTGACAGCTGTCGGCTTTAACTCTTACGCAAT 13232
QY 405 GlyValSerGluSerGlyAlaPro--GlnThrPheThrProAspAlaIleAsnAspLys 423
Db 13233 GGTTCACACCGGCAACCAACCGAAGAACTGACAGCAAGAGCACTCAACCGGGAA 13292
QY 424 ThrGlnGluAlaHisLysGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHisAla 443
Db 13293 ACTCAGACGCGCACTTACAGCGGATTAAGAGCTGATACCGGTGACAAAACCAACCCA 13352
QY 444 SerValValMetTyrSerIleIleAlaSerGluProAlaSerHisGluAspGlyAlaArgGlu 463
Db 13353 AGGGTGTGATGTGGAGATATTGGCAACCAACCGATCCCGTCCGCAAGGTGACCGGAA 13412
QY 464 TyrPheGluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPhe 483
Db 13413 TATTCCCGGCATCTGGCGGAAGCAACCGTAACCTGATCCGACGCTCCGATCACCTGC 13472
QY 484 AlaAsnValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPheAspValSer 503
Db 13473 GTCATGTATATGTTTGCAGCGCTCACACCGATACCATACGATCTTGTATGTGCTG 13532
QY 504 CysIleAsnArgTyrPheGlyTyrTyrSerGlnThrGlyLysPheGluGluAlaGluAla 523
Db 13533 TGCTGAACCGTATTACGGTGTGTATGTCAAAGCGCGATTTGGAACCGCAGAGAAAG 13592
QY 524 AlaLeuGluLysGluLeuHisGlyTyrGlnGluLysPheHisArgProIleValMetThr 543
Db 13593 GTACGAAAAAAGAACTTTCGGCTGCGCAGAGAAACTGCATACGCCGATATCATCAC 13652
QY 544 GlyTyrGlyAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTyrPheSerGlu 563
Db 13653 GAATACGCGCGTGGATACGTAGCCGGGCTGCATCAATGTACACCGCATGTGAGTGA 13712
QY 564 GluPheGlnValGlnMetLeuAspMetTyrHisArgValPheAspArgIleGluSerMet 583
Db 13713 GAAATACAGTGCATGCGTGGATGTATGATCAACCGGCTTTATATGCGCTCAACGCCG 13772
QY 584 AlaGlyLysIleValIleAspPheAlaAspPheGlnThrAsnLeuGlyIleIleArgVal 603
Db 13773 GTGGGTGAACAGTATGAAATTCGCCGATTTGCCACCTCGCAAGGCATATTTGCCGTT 13832
QY 604 AspGlyAsnLysGlyValPheThrArgAspArgLysProLysAlaAlaHisSer 623
Db 13833 GCGCGTAAACAAAGGGGATCTTCAACCGGACCGCAACCGAAGTCGGGGCTTTTCTG 13892
QY 624 LeuArgAlaArgThrPheSerIleAsp 632
Db 13893 CTGCAAAAACGCTGACTGGCATGAAC 13919

RESULT 5
US-10-640-422-160
; Sequence 160, Application US/10640422
; Publication No. US20040229229A1
; GENERAL INFORMATION:

; APPLICANT: Cheo, David
; APPLICANT: Brasch, Michael A.
; APPLICANT: Temple, Gary F.
; APPLICANT: Hartley, James L.
; APPLICANT: Byrd, Devon R.N.
; TITLE OF INVENTION: Use of Multiple Recombination Sites with Unique Specificity in
; TITLE OF INVENTION: Recombinational Cloning
; FILE REFERENCE: 0942.5010004
; CURRENT APPLICATION NUMBER: US/10/640,422
; CURRENT FILING DATE: 2003-08-14
; PRIOR APPLICATION NUMBER: US 60/402,920
; PRIOR FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: US 60/169,983
; PRIOR FILING DATE: 1999-12-10
; PRIOR APPLICATION NUMBER: US 60/188,020
; PRIOR FILING DATE: 2000-03-09
; PRIOR APPLICATION NUMBER: US 09/732,914
; PRIOR FILING DATE: 2000-12-11
; NUMBER OF SEQ ID NOS: 183
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 160
; LENGTH: 5898
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: pms/cw

US-10-640-422-160

Alignment Scores:
Pred. No.: 4,53e-169 Length: 5898
Score: 1715.00 Matches: 338
Percent Similarity: 68.68 Conservative: 94
Best Local Similarity: 53.74 Mismatches: 173
Query Match: 51.13 Indels: 24
DB: 18 Gaps: 11

US-10-757-093-4 (1-634) x US-10-640-422-160 (1-5898)
QY 12 LeuAlaAlaProSerLeuGlyThrProAlaAlaArg--HisPheProArgAsnGluMet 30
Db 1263 TTAAACCATCCCAAGTTTGACAAAAGCAGGCTCATTTAACTTTAAGAAAGATATAT 1322
QY 31 ThrGlnHisGluGlnProLeuIleLysValArgProGlnArgThrSerSerArgGluLeu 50
Db 1323 ACCATG-----GTCGCTCCTGTGAAGAACCCCAACCGCTGAATTC 1361
QY 51 ValAsnLeuAspGlyLeuThrLysPheAlaLeu-----AlaSerGlyLeuAsnAsp 67
Db 1362 AAAAACTGACGCGCTGTGGCATTCAGTTCGATTCGCAAAACCTGTGGAATTGATCAG 1421
QY 68 ThrAlaGlnProTyrThrAlaProLeuProLysGlyLeuGluLysProValProAlaSer 87
Db 1422 -----CGTTGTGGGAAAGCGGCTTACAAAGAACCGGCAATTGCTGTGCCAGCAT 1475
QY 88 TyrAsnAspIlePheIleSerArgLysIleHisAspHisValGlyTyrValTyrTyrGln 107
Db 1476 TTAAAGATCACTTGGCGCATGACATATTCGATTTATGCGCGGCAAGCTGATATCAG 1535
QY 108 ArgGluValIleValProLysGlyTyrPheGlnGluArgTyrLeuValArgAlaGluSer 127
Db 1536 CCGGAAGCTTTATACCGAAGGTTGGCAGGCGCAGCGATATGCTGCTCGTTTCGATCG 1595
QY 128 AlaThrHisLeuGlyAlaArgIleTyrValAsnAspArgLeuValAlaGluHisValGly 147
Db 1596 GTACATCATTTACGGCAAGGTGGGTCAATATTCAGAAAGTGAATGACATCAGGCGGC 1655
QY 148 TyrThrProPheGluAlaAspValThrGluLeuValAlaProGlyGluLysPheArgLeu 167
Db 1656 TATACGCCATTTGAAGCCGATGTCAAGCGGATGTATTATGCGCGGCAAAAGTATGATC 1715
QY 168 ThrIleGlyValAsnAsnGluLeuThrHisGluThrIleProProGlyLysIleThrThr 187
Db 1716 ACCGTTTGTTGAACAAAGAACTGAATCTGCAGACTATCCGCGCGGGAATGTGATTTACC 1775

```



```

FEATURE:
NAME/KEY: promoter
LOCATION: (5557)..(5771)
OTHER INFORMATION: No. US20030175976A1alaine synthetase promoter
FEATURE:
NAME/KEY: misc_feature
LOCATION: (5771)..(5818)
OTHER INFORMATION: MCS multiple cloning site
FEATURE:
NAME/KEY: gene
LOCATION: (5818)..(7717)
OTHER INFORMATION: GUS gene coding for beta glucuronidase
FEATURE:
NAME/KEY: polyA_signal
LOCATION: (7718)..(8447)
OTHER INFORMATION: Poly A from 35S ribosome
FEATURE:
NAME/KEY: misc_feature
LOCATION: (8447)..(8474)
OTHER INFORMATION: MCS multiple cloning site
FEATURE:
NAME/KEY: misc_feature
LOCATION: (8474)..(8647)
OTHER INFORMATION: T-DNA right border
US-09-845-064-11

```

Alignment Scores:

Pred. No.:	8.83e-169	Length:	8654
Score:	1714.50	Matches:	337
Percent Similarity:	69.12%	Conservative:	95
Best Local Similarity:	53.92%	Mismatches:	177
Query Match:	51.12%	Indels:	17
DB:	10	Gaps:	9

US-10-757-093-4 (1-634) x US-09-845-064-11 (1-8654)

```

QY 15 ProSerLeuGlyThrProAlaAlaArgHisPheProArgAsnGluMetThrGlnHisGlu 34
DB 5742 CCNAGGCTTGGCCGCGCTTAACACGCGTGAATCCTT--AATTAAAGTCGCTAGGGGT 5799
QY 35 GlnProLeuIleuValArgProGlnArgThrSerSerArgGluLeuValAsnLeuAsp 54
DB 5800 GGTACGTCCTTATGTTACGTCCTGTAAGAAACCCCAACCCGTAATCAAAAACTCGAC 5859
QY 55 GlyLeuTrpLysPheAlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnPro 71
DB 5860 GGCCTGGGCGCATTCAGTCTCGATCGCAAAACCTGGGAATGATCG-----CGTGG 5913
QY 72 TrpThrAlaProLeuProLysGlyLeuGluCyProValProAlaSerTyraAsnAspIle 91
DB 5914 TGGGAAAGCCCGCTTACAAAGACCCGGCAATTGCTGTGCCAGCGATTTTTAAACGATCAG 5973
QY 92 PheIleSerArgGluIleHisAspHisValGlyTrpValTyrrTrpGlnArgGluValIle 111
DB 5974 TTGCGCGATGCAATATTCGTAATATATGCGGCAAGCTCTGTATACGCGCAAGCTTT 6033
QY 112 ValProLysGlyTrpSerGlnGluArgTyrrLeuValArgAlaGluSerAlaThrHis 131
DB 6034 ATACGAAAGGTGGGAGCGCCAGCGTATCGTGTGCTTCGATGCGGTCACTCATTAC 6093
QY 132 GlyArgIleTyrrValAsnAsnArgLeuValAlaGlnHisValGlyGlyTyrrTrpPhe 151
DB 6094 GGGAAAGTGGGTGCAATATACAGAAAGTATGATGATGATGATGATGATGATGATGATGAT 6153
QY 152 GluAlaAspValThrGluLeuValAlaProGlyGluLysPheArgLeuThrIleGlyVal 171
DB 6154 GAAGCCGATGTCACGCGGTATGTTATGCGCGGAAAAAGTATACGATACCGTTGTGTG 6213
QY 172 AsnAsnGluLeuThrHisGluThrIleProGlyLysIleThrThrArgLysAlaThr 191
DB 6214 AACACAGAACTGAATCGGACACTATCCGCCGGAATGTGATTTCC--GACGAAAC 6270
QY 192 GlyLysArgIleGlnThrTyrrGlnHisAspPheTyraAsnTyrrAlaGlyLeuAlaArgSer 211

```

```

DB 6271 GGCAGAAAAAGACGCTTACTTCATGATTTCTTAACTAATCCGGAATCCATCCAGC 6330
QY 212 IleTrpLeuTyrrSerValProGlnGlnHisIleGlnAspIleThrValThrAspVal 231
DB 6331 GTATGCTCTACACCAAGCCGAAACACCTGGGTGACGATATCACCTGTGTGACGATGTC 6390
QY 232 -----AspGlyAspAsnGlyLeuIleAsnTyrrGluValGlyValAlaAsnGlnThrThr 249
DB 6391 GCGCAGACGTGTAACCAACGCGTCTGTGATGCGACAGT-----GTGCCAAT----- 6438
QY 250 GlyGlnIleGlnIleSerValIleAspGluAspGlyValIleValAlaLysAlaSerGly 269
DB 6439 GGTGATGTACGCGTGAACGCTGATGCGGATCAACAGGTGTTCCAACTGCAACAGGC 6498
QY 270 AlaGlnGlyThrValThrIleProSerValLysLeuTrpGlnProGlyValAlaTyrrLeu 289
DB 6499 ACTAGCGGACCTTGTAGAGTGTGAATCCGCACTTGGCAACCGGTTAAGTTATCTC 6558
QY 290 TyrGlnLeuGlnValAsnIleValGlySerSerGlyAspValValAspThrTyraAsnLeu 309
DB 6559 TATGAATGTGCGTC-----ACAGCCAAAGCCACAGAGTGTATCTACCCGCTT 6612
QY 310 AlaThrGlyValArgThrValLysValAlaGlySerGlnPheLeuIleAsnGlyLysPro 329
DB 6613 CGCGTGGGATCCGTCAGTGGCAGTGGAGGAGGAGCAAGTCTCTGATTAACCAAAACCG 6672
QY 330 PheTyrrPheThrGlyPheGlyLysHisGluAspThrAlaValArgGlyLysGlyHisAsp 349
DB 6673 TTCTACTTTCACGTGGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6732
QY 350 ProAlaTyrrMetValHisAspPheGlnLeuMetLysTyrrIleGlyValaAsnSerPheArg 369
DB 6733 AAGTGTGATGTGTGACGACCAACGATTAATGATGATGATGATGATGATGATGATGATGATGAT 6792
QY 370 ThrSerHisTyrrProTyrrAlaGluGluValMetAspPheAlaAspArgAsnGlyIleVal 389
DB 6793 ACCTCCATTAACCTTACCTTAACCTTAAGATGCTGATGCGGAGATGATGATGATGATGATGAT 6852
QY 390 ValIleAspGlyThrProAlaValGlyLeuAsnIleAlaLeu---MetGlyValSerGlu 408
DB 6853 GTGATGATGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6912
QY 409 SerGlyAlaPro---GlnThrPheThrProAspAlaIleAsnAspLysThrGlnGluAla 427
DB 6913 GGCACAAACCGAAGAAAGTGAACGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 6972
QY 428 HisLysGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHisAlaSerValValMet 447
DB 6973 CACTTACAGCGCATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 7032
QY 448 TrpSerIleAlaAsnGluProLaseHisGlyAspGlyValArgGlyTyrrPheGluPro 467
DB 7033 TGGAGATGTCACAGCAACCGATACCCCTCCGCAAGTGACCGGAAATATTTCCGCGCA 7092
QY 468 LeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGly 487
DB 7093 CTGGCGGAAGCAACCGGTAACTGCAACCGCGCGCATGACCTGCGCATGTAATG 7152
QY 488 ThrAlaThrTyrrGlnLeuAspArgIleSerAspLeuPheAspValSerValIleAsnArg 507
DB 7153 TTCTGCGAGCTCACACCATACCATACCATACCTTTGATGCTGTGCTGTGAACCT 7212
QY 508 TyrPheGlyTyrrTyrrSerGlnThrGlyAspLeuGluGluAlaGluAlaIleLeuGlyLys 527
DB 7213 TATTAAGATGTGATATCCAAAGCCGATTTGGAAACGCGCAAGAAAGTATCTGAAAAA 7272
QY 528 GluLeuHisGlyTyrrGlnGluLysPheHisArgProIleValMetThrGlyTyrrGlyAla 547
DB 7273 GAACCTTCTGCGCTGGAGGAGAACTGATCAGCGGATTAATCATCATCCGAATACGCGCGTG 7332
QY 548 AspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTyrrSerGluGluIleGlnVal 567

```

```

Db 7333 GATACGTTACCGCGGCTGCACTCATATGTACACCGACATGTGAGTGAAGATATCACTGT 7392
Qy 568 GIMetLeuApMeTyrHisArgValPheAspArgIleGluSerMetAlaGlyGluHis 587
Db 7393 GCATGGCTGATGATGATACCGCGGCTTGTGATGCGGTACGCGCGTCTGGCGGAACAG 7452
Qy 588 ValTyrAsnPheAlaAspPheGlnThrAsnLeuGlyIleIleValAlaAspGlyAsnIle 607
Db 7453 GTATGGAATTTCCGCGATTTTTCGACCTCCGACAGCATATTCGCCGTTGGCGGTAAACAG 7512
Qy 608 LyeGlyValPheThrArgAspArgLysProLysAlaAlaHisSerLeuArgAlaArg 627
Db 7513 AAAGGATCTTCACTCCGACCGCAACCGAAGTCGGCGCTTTCTGTGCAAAAAGCGC 7572
Qy 628 TrpThrSerLeuAsp 632
Db 7573 TGGACTGCGATGAAC 7587

RESULT 7
US-09-845-064-14
; Sequence 14, Application US/09845064
; Publication No. US20030175976A1
; GENERAL INFORMATION:
; APPLICANT: MERISTEM THERAPEUTICS
; TITLE OF INVENTION: CLEAN SYNTHETIC VECTORS, PLASMIDS, TRANSGENIC PLANTS
; TITLE OF INVENTION: AND PLANT PARTS CONTAINING SAID VECTORS, AND THEIR
; TITLE OR INVENTION: METHODS OF PRODUCTION
; FILE REFERENCE: SYNVEC1
; CURRENT APPLICATION NUMBER: US/09/845,064
; CURRENT FILING DATE: 2001-04-27
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 8654
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Plasmid
; FEATURE:
; OTHER INFORMATION: pmrT1196
; NAME/KEY: rep origin
; LOCATION: (1)..(654)
; OTHER INFORMATION: Origin of replication ori RK2
; FEATURE:
; NAME/KEY: rep origin
; LOCATION: (655)..(1263)
; OTHER INFORMATION: Origin of replication ori COLEI
; FEATURE:
; NAME/KEY: gene
; LOCATION: (1264)..(2603)
; OTHER INFORMATION: NPT III gene coding for neomycin
; OTHER INFORMATION: phosphotransferase and Kanamycin resistance
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2604)..(4098)
; OTHER INFORMATION: TrfA locus from RK2 coding for two proteins, P2S5
; OTHER INFORMATION: and P382, enabling the increase of the rate of
; OTHER INFORMATION: replication
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4105)..(4271)
; OTHER INFORMATION: T-DNA left border
; FEATURE:
; NAME/KEY: terminator
; LOCATION: (4272)..(4559)
; OTHER INFORMATION: No. US20030175976A1:AlaIine synthetase terminator
; FEATURE:
; NAME/KEY: gene
; LOCATION: (4560)..(5556)
; OTHER INFORMATION: NPT II gene coding for neomycin phosphotransferase
; OTHER INFORMATION: and Kanamycin resistance
; FEATURE:
; NAME/KEY: promoter

```

```

; LOCATION: (5557)..(5771)
; OTHER INFORMATION: No. US20030175976A1:AlaIine synthetase promoter
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (5771)..(5818)
; OTHER INFORMATION: MCS multiple cloning site
; FEATURE:
; NAME/KEY: gene
; LOCATION: (5818)..(7717)
; OTHER INFORMATION: GUS gene coding for beta glucuronidase
; FEATURE:
; NAME/KEY: polyA signal
; LOCATION: (7718)..(8447)
; OTHER INFORMATION: Poly A from 35S ribosome
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (8447)..(8474)
; OTHER INFORMATION: MCS multiple cloning site
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (8474)..(8647)
; OTHER INFORMATION: T-DNA right border
; US-09-845-064-14

Alignment Scores:
Pred. No.: 8,83e-169 Length: 8654
Score: 1714.50 Matches: 337
Percent Similarity: 69.12% Conservative: 95
Best Local Similarity: 53.92% Mismatches: 177
Query Match: 51.12% Indels: 17
DB: 10 Gaps: 9

US-10-757-093-4 (1-634) x US-09-845-064-14 (1-8654)
Qy 15 ProSerLeuGlyThrProAlaAlaArgHisPheProArgLysGluMetThrGlnHisGlu 34
Db 5742 CCGAAGCTGGCGCGCGGCTTAACACCGGTGATCTT--AATTAGGCACTAGAGGCT 5799
Qy 35 GlnProLeuIleLysValAlaProGlnArgThrSerSerArgGluLeuValAsnLeuAsp 54
Db 5800 GGTCACTCCCTTATGTTACGCTCTGTAGAAACCCCAACCGGTAAATCAAAAATCTGCAC 5859
Qy 55 GlyLeuTrpLysPheAlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnPro 71
Db 5860 GGCCTGTGGGCACTTCACTGATCGGAAACCTGCGAATTGATCAG-----CGTTGG 5913
Qy 72 TrpThrAlaProLeuProLysGlyLeuGluCysProValProAlaSerTyrAsnAspIle 91
Db 5914 TGGGAAAGCGGTTACAAAGAACCGGCAATTGCTGTGCGACAGATTTAACGATCAG 5973
Qy 92 PheIleSerArgGluIleHisAspHisValGlyTyrValTyrTyrGlnArgGluValIle 111
Db 5974 TTGCGCGATGCAATATTCCTAATTATGCGGCAAGCTGTGATACGCGCAAGCTTTT 6033
Qy 112 ValProLysGlyTrpSerGlnGluArgTyrLeuValAlaGluSerAlaThrHisHis 131
Db 6034 ATACCGAAGGTTGGGCGACCGCGATGTGCTGTGCTGCTTCCATGCGGCTCACTCATTTAC 6093
Qy 132 GlyArgIleTyrValAsnAsnArgLeuValAlaGluHisValGlyGlyTyrTrpProPhe 151
Db 6094 GGCAAGATGGGTCATATATACGAAGATGATGAGATGAGGCGGCTATACGCCATTTT 6153
Qy 152 GluAlaAspValThrGluLeuValAlaProGlyGlyLysPheAspArgLeuThrIleGlyVal 171
Db 6154 GAAGCCGATGTCAAGCGGTATTTATTTGCGGGAAGAAAGTACGTATACACCGTTGTGTG 6213
Qy 172 AsnAsnGluLeuThrHisGluThrIleProProGlyLysIleThrThrGlyAsnAlaThr 191
Db 6214 AACACGAACTGAACTGCGACACTATTCGCCCGGAATGATGTATTAC--GACGAAAC 6270
Qy 192 GlyLysArgIleGlnThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaArgSer 211
Db 6271 GGCAGAAAGAAAGCAAGCTTACTTCCATGATTTCTTAACTATCCCGAATCATTCGACGC 6330

```

QY 212 ILeuThrLeuValProGlnGlnHisIleGlnAspIleThrValIleAspVal 231
Db 6331 GTATGCTCTACACACGCGGACACCTGGGTGACATATCAACCGGCGGACGATGC 6390
QY 232 -----AspGlyAspAsnGlyLeuIleAsnTrpGluValGluValAlaAsnGlnThr 249
Db 6391 GCCGACAGATGTAACACACGCGTCTGTGACTGGCAGAGTG---GTGGCCAAAT----- 6438
QY 250 GlnGlnIleGlnIleSerValIleAspGluAspGluAlaIleValAlaIleAsnGly 269
Db 6439 GGGATGTCACGGTGAACGTGCGATGCGATCAACAGGTGTGTGAACGTGAACAGGC 6498
QY 270 AlaglnGlyThrValThrIleProSerValIleSerTrpGlnProGlnValAlaIle 289
Db 6499 ACAGACGCGACTTTGCAACGTGATCCGACACTTCGACACCGGTTGAAGTTATCTC 6558
QY 290 TyrGlnIleGlnValAlaAsnIleValGlySerSerGlyAspValValAspThrTyrAsnLeu 309
Db 6559 TATGAACGTGCGTC-----ACAGCCAAAGCCAGACAGATGTGATCTACCGCTT 6612
QY 310 AlatThrGlyValArgThrValIleValAlaGlySerGlnPheLeuIleAsnGlyIlePro 329
Db 6613 CGCGTGGCGATCCGCTGAGTGAAGGAGGCGAAGAGTCTGTGATTAACAACACG 6672
QY 330 PheTyrPheThrGlyPheGlyIleGlyAspThrAlaValArgGlyIleGlyIleAsp 349
Db 6673 TTCTACTTACTGCTTGTGCTGCTGATGAAGATCCGACACTTGGCTGGCAAGATTCGAT 6732
QY 350 ProAlaTyrMetValHisAspPheGlnLeuMetIleTrpIleGlyAlaAsnSerPheArg 369
Db 6733 AACGTGCTGATGGTGCACGACGACGATTAATGACGATGGTGGCGCAACTCTACCGT 6792
QY 370 ThrSerHisTyrProTyrAlaGlnGluValMetAspPheAlaAspArgAsnGlyIleVal 389
Db 6793 ACCTGCGATTAACCTTAACGCTGAAGATGCTGACGCGGACATGAACATGCGATCGT 6852
QY 390 ValIleAspGlnThrProAlaValGlyLeuAsnIleAlaLeu---MetGlyValSerGlu 408
Db 6853 GTGATGATGAACACTGCTGCTGCTTAACTCTCTTAACTGTCGATTTGTTGGAAGCG 6912
QY 409 SerGlyAlaPro---GlnThrPheThrProAspAlaIleAsnAspIleThrGlnGluAla 427
Db 6913 GGCACACGCGGACGAAAGCTGACGAGGAAAGCGAGTCAACGCGGAAATCTCAGCAAGCG 6972
QY 428 HisIleGlnAlaIleAsnGlnLeuIleAlaArgAspIleAsnHisAlaSerValIleMet 447
Db 6973 CACTTACAGCGGATTAAGAGCTGATGCGCGTGAACAAACCAACCCAGCGTGTGATG 7032
QY 448 TrpSerIleAlaAsnGlnProAlaSerHisGlnAspGlyAlaArgGluTyrPheGluPro 467
Db 7033 TGGAGTATTGSCAACGACGAGATACCCGTCGCAAGGTGACAGGAGATTTTCGCGCCA 7092
QY 468 LeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGly 487
Db 7093 CTGGCGGAAACGACGCGTAACTGACCCGACCGCGTCCGATTCACCTGCGCAATGTAATG 7152
QY 488 ThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArg 507
Db 7153 TTCTGCGACGCTACACCGATACCATGAGATCTCTTGAATGTGCTGTCGTAACCGT 7212
QY 508 TyrPheGlyTyrPyrSerGlnThrGlyAspLeuGlnGluAlaIleAlaLeuGlnIleVal 527
Db 7213 TATTACGAGATGATGTCGAAGCGGATTTGGAAACGCGACAGAAAGATCTGGAAGAAA 7272
QY 528 GluLeuHisGlyTyrGlnGluIlePheHisArgProIleValMetThrGluTyrGlyAla 547
Db 7273 GAACTTCTGGCTGGCGGAGGAAATGCACTGACCGCAATTCATCAACCGAATCGCGCGTG 7332
QY 548 AspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTyrSerGlnGluIlePheGlnVal 567
Db 7333 GATACGTTAGCCGGGCTGACCTCATATGTACCGACGACATGTGAGTGAAGATTCAGTGT 7392

QY 568 GlnMetLeuAspMetTyrHisArgValPheAspArgIleGluSerMetAlaGlyGlnHis 587
Db 7393 GCATGCTGGATATGATATCACCGCGTCTTGAATCGGTGACGCGCGTCCGTGGAACG 7452
QY 588 ValTrpAsnPheAlaAspPheGlnThrAsnLeuGlyIleIleArgValAspGlyAsnIle 607
Db 7453 GTATGGAATTTCCCGCATTTTGGCACTGCAAGGCGATTTGGCGCTTGGCGGTAAACAAG 7512
QY 608 IysGlyValPheThrArgAspArgIleProIleValAlaAlaHisSerLeuArgAlaArg 627
Db 7513 AAAGGATCTTCACTGCGACGCGAACCAGAACTGCGCGCTTTTCTGCTGCAAAAAACG 7572
QY 628 TrpThrSerIleAsp 632
Db 7573 TGGACTGGCATGAC 7587
RESULT 8
US-09-845-064-22
Sequence 22, Application US/09845064
Publication No. US20030175976A1
GENERAL INFORMATION:
APPLICANT: MERISTEM THERAPEUTICS
TITLE OF INVENTION: CLEAN SYNTHETIC VECTORS, PLASMIDS, TRANSGENIC PLANTS
TITLE OF INVENTION: AND PLANT PARTS CONTAINING SAID VECTORS, AND THEIR
TITLE OF INVENTION: METHODS OF PRODUCTION
FILE REFERENCE: Synvec1
CURRENT APPLICATION NUMBER: US/09/845,064
CURRENT FILING DATE: 2001-04-27
NUMBER OF SEQ ID NOS: 57
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 22
LENGTH: 8987
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Plasmid
FEATURE:
OTHER INFORMATION: pmr1212
NAME/KEY: rep origin
LOCATION: (1)..(654)
OTHER INFORMATION: Origin of replication ori RK2
FEATURE:
NAME/KEY: rep origin
LOCATION: (1264)..(2603)
OTHER INFORMATION: NPT III gene coding for neomycin transferase and
OTHER INFORMATION: kanamycin resistance
FEATURE:
NAME/KEY: misc feature
LOCATION: (2604)..(4098)
OTHER INFORMATION: TrfA locus from RK2 coding for two proteins, P285
OTHER INFORMATION: and P382, enabling the increase of the rate of
OTHER INFORMATION: replication
FEATURE:
NAME/KEY: gene
LOCATION: (1264)..(2603)
OTHER INFORMATION: NPT III gene coding for neomycin transferase and
OTHER INFORMATION: kanamycin resistance
FEATURE:
NAME/KEY: terminator
LOCATION: (4272)..(4559)
OTHER INFORMATION: No. US20030175976A1AlaIle synthetase terminator
FEATURE:
NAME/KEY: gene
LOCATION: (4575)..(5150)
OTHER INFORMATION: Bar gene coding for phosphinothricine
OTHER INFORMATION: acetyltransferase and glutosinate resistance
FEATURE:
NAME/KEY: promoter
LOCATION: (5151)..(5368)


```

Db 7666 GATAGTAGCCGGCTGCATCATGTACACCCAGTGTGAGTGAAGATATCAGTGT 7725
Qy 568 GlnMetLeuAspMetTyrHisArgValPheAspArgIleGluSerMetAlaGlyGluHis 587
Db 7726 GCATGCTGGATGATATATACACCGCGCTTTTATCCCGCCAGCGCTGTCGTGGTGAACAG 7785
Qy 588 ValTTrpAspPheAlaAspPheGlnThrAsnLeuGlyIleIleArgValAspGlyAsnLys 607
Db 7786 GTATGATTTTGGCGATTTTCCGATCTCGCAGGCAATTTGCGGCTTGGCGGTAAACAG 7845
Qy 608 LysGlyValPheThrArgAspArgIleProLysAlaAlaHisSerLeuArgAlaArg 627
Db 7846 AAAGGATCTTCACCTCGCGACCGCAACCGAAGTCGCGCTTTCTGCTCAAAAAGC 7905
Qy 628 TrpThrSerIleAsp 632
Db 7906 TGGACTGGCATGAAC 7920

RESULT 9
US-09-845-064-18
: Sequence 18, Application US/09845064
: Publication No. US20030175976A1
: GENERAL INFORMATION:
: APPLICANT: MERISTEM THERAPEUTICS
: TITLE OF INVENTION: CLEAN SYNTHETIC VECTORS, PLASMIDS, TRANSGENIC PLANTS
: TITLE OF INVENTION: AND PLANT PARTS CONTAINING SAID VECTORS, AND THEIR
: FILE REFERENCE: SynVec1
: CURRENT APPLICATION NUMBER: US/09/845,064
: NUMBER OF FILING DATE: 2001-04-27
: NUMBER OF SEQ ID NOS: 57
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 18
: LENGTH: 9390
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Plasmid
: FEATURE:
: OTHER INFORMATION: PMRT1204
: NAME/KEY: rep origin
: LOCATION: (1)-(654)
: OTHER INFORMATION: Origin of replication ori RK2
: FEATURE:
: NAME/KEY: rep origin
: LOCATION: (655)-(1263)
: OTHER INFORMATION: Origin of replication ori COLE1
: FEATURE:
: NAME/KEY: gene
: LOCATION: (1264)-(2603)
: OTHER INFORMATION: NPT III gene coding for neomycin
: OTHER INFORMATION: phosphotransferase and kanamycin resistance
: FEATURE:
: NAME/KEY: misc feature
: LOCATION: (2604)-(4098)
: OTHER INFORMATION: TrfA from RK2 coding for two proteins, P285 and
: OTHER INFORMATION: P382, enabling the increase in the rate of
: FEATURE:
: NAME/KEY: misc feature
: LOCATION: (4105)-(4271)
: OTHER INFORMATION: T-DNA left border
: FEATURE:
: NAME/KEY: terminator
: LOCATION: (4272)-(4559)
: OTHER INFORMATION: No. US20030175976A1:aline synthetase terminator
: FEATURE:
: NAME/KEY: gene
: LOCATION: (4560)-(5559)
: OTHER INFORMATION: Wild type NPT II gene coding for neomycin
: OTHER INFORMATION: phosphotransferase and kanamycin resistance
: FEATURE:

```

```

: NAME/KEY: promoter
: LOCATION: (5560)-(5771)
: OTHER INFORMATION: No. US20030175976A1:aline synthetase promoter
: FEATURE:
: NAME/KEY: promoter
: LOCATION: (5772)-(6514)
: OTHER INFORMATION: Enhanced promoter from 35S ribosome
: FEATURE:
: NAME/KEY: misc feature
: LOCATION: (6514)-(6554)
: OTHER INFORMATION: MCS multiple cloning site
: FEATURE:
: NAME/KEY: gene
: LOCATION: (6554)-(8453)
: OTHER INFORMATION: GUS gene coding for beta glucuronidase
: FEATURE:
: NAME/KEY: polyA signal
: LOCATION: (8454)-(9183)
: OTHER INFORMATION: Poly A from 35S ribosome
: FEATURE:
: NAME/KEY: misc feature
: LOCATION: (9183)-(9210)
: OTHER INFORMATION: MCS multiple cloning site
: FEATURE:
: NAME/KEY: misc feature
: LOCATION: (9210)-(9383)
: OTHER INFORMATION: T-DNA right border
: US-09-845-064-18

Alignment Scores:
Pred. No.: 9,92e-169 Length: 9390
Score: 1714.50 Matches: 337
Percent Similarity: 69.12% Conservative: 95
Best Local Similarity: 53.92% Mismatches: 177
Query Match: 51.12% Indels: 17
DB: Gaps: 9

US-10-757-093-4 (1-634) x US-09-845-064-18 (1-9390)
Qy 15 ProSerLeuGlyThrProAlaAlaArgHisPheProAlaGlnMetThrGlnIleGlu 34
Db 6478 CCCAAGCTTGGCGCGCGCGCTTACACCGCGTGAATCCTT--ATTGAAGTCACTTGAAGGT 6535
Qy 35 GlnProLeuIleValArgProGlnArgThrSerArgGluLeuValAsnLeuAsp 54
Db 6536 GGTCAAGTCCCTTATGATGTCCTTGAAGAACCCCAACCGGTGAATCAAAAACCTGAC 6595
Qy 55 GlyLeuTrpLysPheAlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnPro 71
Db 6596 GGCCTGTGGCATTCAGTCTGATCGCAAAAAGTGTGAATTGATCAG-----CGTTGG 6649
Qy 72 TrpThrAlaProLeuProLysGlyLeuGlyProValProAlaSerTyrAsnAspIle 91
Db 6650 TGGAAAGCGCGCTTCAAGAAAGCCCGGCAATGCTGTGCCAGGAGATTTAACGATCAG 6709
Qy 92 PheIleSerArgGluIleHisAspHisValGlyTrpValTyrTyrGlnArgGluValIle 111
Db 6710 TTCGCCGATGCGATATTCGTAATATGCGGCAACGCTGATTCAGCGGAAGTCTT 6769
Qy 112 ValProLysGlyTrpSerGlnArgTyrLeuValArgAlaGluSerAlaThrHisHis 131
Db 6770 ATACCGAAGGTTGGCGAGCGCAGGATCGTGTGCTTCGATCGCGGTACATCATTAC 6829
Qy 132 GlyArgIleTyrValAsnAsnArgLeuValAlaGluHisValGlyTyrThrProPhe 151
Db 6830 GCGAAAGTGTGGCTCAATATCAGGAAGTGAAGATGAGGCGGCTTATACCCATT 6889
Qy 152 GluAlaAspValThrGluLeuValAlaProGlyGluLysPheArgLeuThrIleGlyVal 171
Db 6890 GAAGCCGATGTCACGCGGTATGATTCGCGGAAAGTGAAGTACATACCGCTTGTGTG 6949
Qy 172 AsnArgIleLeuThrHisGluThrIleProProGlyLysIleThrThrGlyAsnAlaThr 191

```

Dh	6550	AACAAGCAATCGAATCGCAGACATATCCCGCGGGAATGGATTAAC---GACGAAAC	7006
Oy	192	GLYLYAAGLIEGLINTHTYGLINHIASPhETyRANTYRAAGLYLEUALAASer	211
Dh	7007	GGCAGAGAAAAGCAGTCTTACTTCATATTTCTTAACTTGGCCGATCATTCGCGAC	7066
Oy	212	ILetRpleuTYRserValProGlnGlnHISILEGIAAspLIEthrValIThrAspVal	231
Dh	7067	GTAATGCTTACACACCGCCCAACCTCGGGTGAACGATATCACCGTGTATCCGATCTC	7126
Oy	232	-----AspGLYAspAsnGLYLeuILEAsnTYRGLValGLUValLLeAsnGLNThrThr	249
Dh	7127	GGCGAAGCTGTAACACGACCGCTGTGATCGGACGGT---GTGGCCAAAT-----	7174
Oy	250	GLYGLNILEGNIIESerValILEAspGLUAspGLYAlALEValLALeValIASerGLY	269
Dh	7175	GGTGTATGCACCGTTGAACTGCGTATCCGATCAACAGGAGTGTTCACACTGCAAAAGGC	7234
Oy	270	ALAAGNGLYThrValIThrLEProserValLYsLeuThrPGINProGLYAlALATYRLeu	289
Dh	7235	ACTAGCGGAGACTTTCGAAGTGGTGAATCCGACCTCTGGCAACGGGAGAAAGTTATCTC	7294
Oy	290	TYRGLNLeuGlnValAsnLILEValGLYserSerGLYAspValValAspThrTYRAsnLeu	309
Dh	7295	TATGAACCTGTGGTC-----ACAGCCAAAAGCCAGACAGTGTGATATCTACCCGCTT	7348
Oy	310	ALAThrGLYValARgThrValLYValALAGLYserGlnPheLeuLLeAsnGLYLeuPro	329
Dh	7349	CGCGTCGGCATCCCGGTCAGTGGCAGTGAAGGGCCAAAGTTCTCGATTAACCAAAACCG	7408
Oy	330	PheTYRPhEThrGLYPhGGLYLeHISGLUAspThrLALeValARGLYLYGLYHISAsp	349
Dh	7409	TTCTACTTACTCGCTGGTGTGTGTCATGAATGCGGACCTTGGCGAAAGGATTCAT	7468
Oy	350	ProALATYRmetValHISAspPheGlnLeuMetLYSTRILEGLYALeAsnSerPheArg	369
Dh	7469	AACGGCGTGAATGGTCAACGACCAACGCAATTAATGACTGATTTGGGCCAACTCTCAACGT	7528
Oy	370	ThISerHISLYRProTYRAlAGLUValMeAspPheALeAspARgAsnGLYLEVal	389
Dh	7529	ACCTGCGCATTAACCTTACGCTGAAGAGATGCTCGACTGGCAGATGAACGAGCATCGTG	7588
Oy	390	ValLLeAspGLUThrProAlValGLYLeuAsnLLeAlLeu---MeGLYValSerGLU	408
Dh	7589	GTGATGTATGAACCTGCTGTGCGCTTTAACTCTTTAAGCATTTGGTTCGAACGG	7648
Oy	409	SerGLYAlAPro---GlnThrPheThrProAspAlALEAsnAspLYThrGlnGLUVal	427
Dh	7649	GGCAACAAGCCGAAAGAACTGTACACGCAAGACGACGTCMAACGGGGAACCTCAGCAACGG	7708
Oy	428	HISLYSGLNALALEArgLLeuLILEAlARgAspLYSAsnHISAlASerValValMet	447
Dh	7709	CACCTTACAGGGGATTAAGAGCTGATACGGGTGACAAAAACCAACCAAGGTGTGTATG	7768
Oy	448	TPSerLILEAlAsnGLUProLAserHISGLUAspGLYAlARgGLUTYRPhEGLUPro	467
Dh	7769	TGGAGTATTTGCCAAGCAACGAGATCCCGTCCGCAAGGTGACAGGGATATTTCGGCCCA	7828
Oy	468	LeuThrAsnLeuThrARGLNLeuAspProThrARgProLIEthrPheAlAsnValGLY	487
Dh	7829	CTGGCGGAGCAACCGGTAACTGACCCGACGCGCTCCGATCACTGCGTCAAAATGTATG	7888
Oy	488	ThrAlAThrTYRGLNLeuAspARgLIESerAspLeuPheAspValserCYsLLeAsnArg	507
Dh	7889	TTTCGCCAGCGTCACACCGATACATACGATCTCTTGATGTGCTGTGCTGAACCGT	7948
Oy	508	TYRPhEGLYTYRYSerGlnThrGLYAspLEUGLValAGLUAlALeUGLLeuLYS	527
Dh	7949	TATTACGATGTGATGTCCAAAGCGCGCATTTGGAAACGGCAGAGAAAGTCTGGAAAAA	8008
Oy	528	GLIUEuHISGLYTRPGLNGLYLeuPheHISARgProLIEValMetThrGLUTYRGLYAlA	547
Dh	8009	GAACCTCTGGCGCTGCAGAGAAATGTGATCAACCGGATTAATCATCAACGGAATACGGCGTG	8068

```

Oy 548 AsphthleuMlaGlyLeuH1sSer1leuCluYleuProTTPSerGIuGluPhaG1Val 567
Db 8069 GATACGTTAGCCCGGAGCTGCACCTCAATGTACACCGACATGTGAGTGAAGATACACTGT 8128
Oy 568 G1metLeuAmpMetCTYrh1sArgVal1PheAAsphrg1leuGserMetAlaGlyGluH1s 587
Db 8129 GCATGGCTGGATATGTATGTACACCGCGCTTTTGATCGCGTACAGCGCGTGTGAACAG 8188
Oy 588 ValTTPAsnPhaAlaAAspPheG1nThAsnleuCluYle1leAArgValAAspG1YAsnLys 607
Db 8189 GTATGGAAATTTCCGCGATTTCGACCTCGCAAGGCATATTGCGCGTTGCGGTPAACAG 8248
Oy 608 LysG1YValPheTrrrAAspArgLysProGlyAlaAlaH1sSer1leuAArgAlaArg 627
Db 8249 AAAGGATCTTCACCTCCGACCGAACCAGAACTCGCGCTTTCTGCTGCAAAAACG 8308
Oy 628 TrpThSer1leAep 632
Db 8309 TGGACTGGCATGAAC 8323

RESULT 10
US-09-845-064-20
/ Sequence 20, Application US/09845064
/ Publication No. US20030175976A1
/ GENERAL INFORMATION:
/ APPLICANT: MERISTEM THERAPEUTICS
/ TITLE OF INVENTION: CLEAN SYNTHETIC VECTORS, PLASMIDS, TRANSGENIC PLANTS
/ TITLE OF INVENTION: AND PLANT PARTS CONTAINING SAID VECTORS, AND THEIR
/ FILE REFERENCE: SynVec1
/ CURRENT APPLICATION NUMBER: US/09/845, 064
/ CURRENT FILING DATE: 2001-04-27
/ NUMBER OF SEQ ID NOS: 57
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 20
/ LENGTH: 9390
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence:Plasmid
/ OTHER INFORMATION: pmr11206
/ FEATURE:
/ NAME/KEY: rep_origin
/ LOCATION: (1)..(654)
/ OTHER INFORMATION: Origin of replication ori RK2
/ FEATURE:
/ NAME/KEY: rep_origin
/ LOCATION: (655)..(1263)
/ OTHER INFORMATION: Origin of replication ori COLE1
/ FEATURE:
/ NAME/KEY: gene
/ LOCATION: (1264)..(2603)
/ OTHER INFORMATION: NPT III gene coding for neomycin
/ OTHER INFORMATION: phosphotransferase and kanamycin resistance
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (2604)..(4098)
/ OTHER INFORMATION: trfA locus from RK2 coding for two proteins, P265
/ OTHER INFORMATION: and P382, enabling the increase of the rate of
/ OTHER INFORMATION: replication
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (4106)..(4271)
/ OTHER INFORMATION: T-DNA left border
/ FEATURE:
/ NAME/KEY: terminator
/ LOCATION: (4272)..(4559)
/ OTHER INFORMATION: No. US20030175976A1alaine synthetase terminator
/ FEATURE:
/ NAME/KEY: gene
/ LOCATION: (4560)..(5559)
/ OTHER INFORMATION: NPT II gene coding for neomycin phosphotransferase

```

```

OTHER INFORMATION: and kanamycin resistance
FEATURE:
NAME/KEY: promoter
LOCATION: (5560)..(5771)
OTHER INFORMATION: No. US20030175976A1aline synthetase promoter
FEATURE:
NAME/KEY: promoter
LOCATION: (5772)..(6514)
OTHER INFORMATION: Enhanced promoter from 35S ribosome
FEATURE:
NAME/KEY: misc feature
LOCATION: (6514)..(6554)
OTHER INFORMATION: MCS multiple cloning site
FEATURE:
NAME/KEY: gene
LOCATION: (6554)..(8453)
OTHER INFORMATION: GUS gene coding for beta glucuronidase
FEATURE:
NAME/KEY: polyA_signal
LOCATION: (8454)..(9183)
OTHER INFORMATION: Poly A from 35S ribosome
FEATURE:
NAME/KEY: misc feature
LOCATION: (9183)..(9210)
OTHER INFORMATION: MCS multiple cloning site
FEATURE:
NAME/KEY: misc feature
LOCATION: (9210)..(9383)
OTHER INFORMATION: T-DNA right border
US-09-845-064-20

```

```

Alignment Scores:
Pred. No.: 9.92e-169 Length: 9390
Score: 1714.50 Matches: 337
Percent Similarity: 69.12% Conservative: 95
Best Local Similarity: 53.92% Mismatches: 177
Query Match: 51.12% Indels: 17
DB: Gaps: 9

```

US-10-757-093-4 (1-634) x US-09-845-064-20 (1-9390)

```

QY 15 ProSerLeuGlyThrProAlaAlaArgHisPheProArgSngluMetThrGlnHisGlu 34
Db 6478 CCCAGCTTGCGCGCGCGGCTTAAACAGCGTGGATCCTT-AAATTAAAGCTCACTAGCGGT 6535
QY 35 GlnProLeuIleuValArgProGlnArgThrSerSerArgGluLeuValAsnLeuAsp 54
Db 6536 GGTCAAGTCCCTTATGTACGTCCTGTAGAAACCCCAACCCGCTGAATAAATAAATCGAC 6595
QY 55 GlyLeuTrpIlePheAlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnPro 71
Db 6596 GGCCTTGCGCGCATTCAGTCTGATCGCGAAACCTGTGAATTGATCG-----CGTTGG 6649
QY 72 TrpThrAlaProLeuProLysGlyLeuGluCysProValProAlaSerTyraAsnAspIle 91
Db 6650 TGGGAAAGCGCGTTACAAAGAAAGCGCGCAATTCCTGCTGCCAGGCACTTTTAAGACAG 6709
QY 92 PheIleSerArgGluIleHisAspHisValGlyTrpValTyryGlnArgGluValIle 111
Db 6710 TTGGCCGATGAGATATTCGAATTATGCGGCAACCTGTGATCGACGCGCAAGTCTTT 6769
QY 112 ValProLysGlyTrpSerGlnGluArgTyryLeuValArgAlaGluSerAlaThrHisHis 131
Db 6770 ATACCGAAAGGTGGGAGCGCGCATCTGCTGTGCTTCCATGCGGTCACTCATTAAC 6829
QY 132 GlyArgIleTyryValAsnAspArgLeuValAlaGluHisValGlyGlyTyryThrProPhe 151
Db 6830 GCGCAAGTGTGCGTCATTAACAGAAAGTGAATGAGATCAGCGCGCTTAAGCCATT 6889
QY 152 GlnAlaAspValThrGluLeuValAlaProGlyGlySphaArgLeuThrIleGlyVal 171
Db 6890 GAAGCCATGTCAAGCGGTATGTTATTCGCGGAAAGTGAATGATACACCGTTTGTGTG 6949

```

```

QY 172 AsnAsnGluLeuThrHisGlyThrIleProProGlyGlySleThrThrGlyAsnAlaThr 191
Db 6950 AACAAAGAACTGAACGTGACAGACTATCCCGCGGAAATGATGATACC---GACAAAAAC 7006
QY 192 GlyValArgIleGlnThrTyryGlnHisAspPheTyraSerTyryValGlyLeuAlaArgSer 211
Db 7007 GCGCAAGAAAGCACTTACTTATTCATGATTTCTTTAATATGCGCGAATCATCATCGACAGC 7066
QY 212 IleTrpLeuTyrySerValProGlnGlnHisIleGlnAspIleThrValValThrAspVal 231
Db 7067 GTAAATGCTTACACACCGCGCAACCTGGGAGCATATTCACCGTGTGACGATGTC 7126
QY 232 -----AspGlyAspAsnGlyLeuIleAsnTyryGluValAlaAsnGlnThrThr 249
Db 7127 GCGCAAGCTGTAAACACAGCGCTGTGATCGGACGTC---GTGGCCAT----- 7174
QY 250 GlyGlnIleGlnIleSerValIleAspGluAspGlyValAlaIleValAlaIleAsnGly 269
Db 7175 GGTGATGTCAAGCGTTGAACGTGATCGGATCAACAGTGTGTGCAACTGACAGAGGC 7234
QY 270 AlaGlnGlyThrValThrIleProSerValLysLeuTrpGlnProGlyValAlaIleThrLeu 289
Db 7235 ACTAGCGGACCTTTCAGATGTGAATCGCACCTCTGGCAACCGGTGAACGTTATCTC 7294
QY 290 TyryGlnLeuGlnValAsnIleValGlySerSerGlyAspValValAspThrTyryAsnLeu 309
Db 7295 TATGAACGTGCGTC-----ACAGCCAAAGCCAGACAGATGTGATATCAACCGGCTT 7348
QY 310 AlaThrGlyValAlaArgThrValLysValAlaGlySerGlnPheLeuIleAsnGlyLysPro 329
Db 7349 CCGGTGCGCATCCGGTCAAGTGCAGTGAAGGCGCAACGTTCTTATTAACCAACAAACCG 7408
QY 330 PheTyryPheThrGlyPheGlyLysHisGlyAspThrAlaValAlaArgGlyLysHisAsp 349
Db 7409 TCTTACTTATTCAGCTTGTGTGTGTGATGAAGATGCGGACTTCTGTGGCAAGATTCGAT 7468
QY 350 ProAlaTyryMetValHisAspPheGlnLeuMetLysTrpIleGlyValAlaAsnSerPheArg 369
Db 7469 AACGTGCTGATGTGTGACACACACGATTAATGATGATTTGGGGCAATCTTACCGCT 7528
QY 370 ThrSerHisTyryProTyryAlaGluGluValMetAspPheAlaAspArgAsnGlyIleVal 389
Db 7529 ACTCGCATTAACCTTACCTGACAGATGATCGTCTGCGCAAGATGATGATGATGATCGG 7588
QY 390 ValIleAspGlnThrProAlaValGlyLeuAsnIleAlaLeu---MetGlyValSerGlu 408
Db 7589 GTGATGATGAAGATGCTGCTGCTGCTTAACTCTCTTTAGCGATGTGTTCCAGAGCG 7648
QY 409 SerGlyAlaPro---GlnThrPheThrProAspAlaIleAsnAspLysThrGlnAla 427
Db 7649 GCGCAACACCCAAAGAAAGTGAACAGCAAGAGGCGATCAACGGGAAATCAGCAAGCG 7708
QY 428 HisLeuGlnAlaIleArgGluLeuIleAlaGlyAspLysAsnHisAlaSerValValMet 447
Db 7709 CACTTACAGGCGATTAAGAGCTGATAGCGCGTGAACAAACCAACCAACCGCTGTGATG 7768
QY 448 TrpSerIleAlaAsnGluProAlaSerHisGlyAspGlyValArgGlyTyryPheGluPro 467
Db 7769 TGGAGTATGTCACCAACACCGGATACCCCTCGCAAGGTGACAGGAAATTTCCGCCCA 7828
QY 468 LeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGly 487
Db 7829 CTGGCGGAAGCAACGGTAAATCGACCGCGATCGCATCGCTGATGATGATG 7888
QY 488 ThrAlaThrTyryGlnLeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArg 507
Db 7889 TTCTGACAGCTCACACCATATCAACAGCATCTTGTGATGTGTGTGCTGAAACCGT 7948
QY 508 TyryPheGlyTrpTyrySerGlnThrGlyAspLeuGluValAlaGluAlaIleLeuGlyLys 527
Db 7949 TATTTAGAGAGGATGTCCAAAGCGCGATTTGAAACCGCAAGAAAGATCTCGAAAAA 8008
QY 528 GlyLeuHisGlyTyryGlnGluLysPheHisArgProIleValMetThrGluTyryAla 547

```

```

|||||
Db      8009 GAACCTTCGCGCCGACGAGGAAACCTGCATCAGCCGATTATCATCACCGAATACGCGCGTG 8068
Qy      548 AspThrLeuAlaGlyLeuHisSerLleuGlyLeuProTyrPheSerGluGluPheGlnVal 567
Db      8069 GATACGTTAGCCGGCGCTGCATCATGATCAACCCAGCATGTGTGAAGAGTATCATGCTGT 8128
Qy      568 GlnMetLeuAspMetTyrHisArgValPheAspArgLleGluSerMetAlaGlyLysHis 587
Db      8129 GCATCGCTGATATGATATCATCCGCGCTTTCATTCGCGCTCAGCCGCGTCTGCGTGAACAG 8188
Qy      588 ValTyrAsnPheAlaAspPheGlnThrAsnLeuGlyLleLeuArgValAspGlyAsnLys 607
Db      8189 GTATGGAATTTCCGCGCATTTTGCACCTCCGACGACGATATTGCGGTTGGCGGTAAACAG 8248
Qy      608 LysGlyValPheMetTyrAspArgValPheProLysValAlaAlaHisSerLeuArgAlaArg 627
Db      8249 AAGGAGATCTTCATCTCCGACCGCAACCGAAGTGGCGGCTTTTCTGCTGCAGAAAAGCG 8308
Qy      628 TrpThrSerLleAsp 632
Db      8309 TGGACTGCGCATGAAC 8323

```

```

RESULT 11
US-09-845-064-50
; Sequence 50, Application US/09845064
; Publication No. US20030175976A1
; GENERAL INFORMATION:
; APPLICANT: MERISTEM THERAPEUTICS
; TITLE OF INVENTION: CLEAN SYNTHETIC VECTORS, PLASMIDS, TRANSGENIC PLANTS
; TITLE OF INVENTION: AND PLANT PARTS CONTAINING SAID VECTORS, AND THEIR
; FILE REFERENCE: SynVec1
; CURRENT APPLICATION NUMBER: US/09/845,064
; CURRENT FILING DATE: 2001-04-27
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 50
; LENGTH: 9688
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: pmRT1334
; NAME/KEY: misc_feature
; LOCATION: (1)
; OTHER INFORMATION: pmRT1334 was obtained by replacing the mpRII
; OTHER INFORMATION: expression cassette of pmRT1206 by the mpRII
; OTHER INFORMATION: expression cassette of pBIN19
US-09-845-064-50

```

```

Alignment Scores:
Pred. No.: 1.04e-168 Length: 9688
Score: 1714.50 Matches: 337
Percent Similarity: 69.12% Conservative: 95
Best Local Similarity: 53.92% Mismatches: 177
Query Match: 51.12% Indels: 17
DB: 10 Gaps: 9

```

```

US-10-757-093-4 (1-634) x US-09-845-064-50 (1-9688)
Qy      15 ProSerIleuThrProAlaAlaArgHisPheProArgAsnGluMetThrGlnHisGlu 34
Db      6776 CCCAAGCTTGGCCGCGCGCTTAACACGCGTGAGTCTT--AAATTAAGTCAGACTAGGGGT 6833
Qy      35 GlnProLeuIleuValArgProGlnArgThrSerSerArgGluLeuValAsnLeuAsp 54
Db      6834 GGTCAAGTCCCTTAAGTACGTCCTGTAGAAACCCCAACCCGCGTGAATCAAAAAAAGTCAAG 6893
Qy      55 GlyLeuTrpLysPheAlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnPro 71
Db      6894 GGCCTGTGGCAATTCAGTCTGGATCGGAAAACCTGTGAATGTATGATCAG-----CGTTGG 6947

```

```

Qy      72 TrpThrAlaProLeuProLysGlyLeuGluCysProValProAlaSerTyrAsnAspIle 91
Db      6948 TGGGAAAGCGCGTTACAAAGAAACCGGGCAATTGCTGTGCGACGACGTTTAAACGATCAG 7007
Qy      92 PheIleSerArgGluLleHisAspHisValGlyTrpValTyrTyrGlnArgGluValIle 111
Db      7008 TTCGCCGATCCAAATATTCCTAAATTATGCGGGCAACCTCTCGATATCACCGCCGAAGTCTTT 7067
Qy      112 ValProLysGlyThrPheSerGlnArgTyrLeuValArgAlaGluSerAlaThrHisHis 131
Db      7068 ATACCGAAAGGTTGGGCGAGCGCATGCTGCTGCTGCTTCATTCGATCGCTCATCTATAC 7127
Qy      132 GlyArgLleTyrValAsnAsnArgLeuValAlaGluHisValGlyGlyTyrThrProPhe 151
Db      7128 GCGAAAGTGGGTCATATATCAAGAGTATGAGAGATCAGGCGCGCTATACGCCATTT 7187
Qy      152 GluAlaAspValThrGluLeuValAlaProGlyGluLysPheAspArgLeuThrLleGlyVal 171
Db      7188 GAAGCCGATGTCAAGCCGCTATGTTATTTGCGGGAAAAGTGTACGATACCGCTTGTGTG 7247
Qy      172 AsnAsnGluLeuThrHisGluThrLleProProGlyLysLleThrThrGlyAsnAlaThr 191
Db      7248 AACCAACGAACTGAACCTGGCAGACTATCCCGCGGAATGGTGAATTACC--GACGAAAC 7304
Qy      192 GlyLysArgLleGlnThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaArgSer 211
Db      7305 GCGAAGAAAGAGAGTCTTACATTCATGATTTCTTAACTATCCGGAATCATCCGAC 7364
Qy      212 IleTrpLeuTyrSerValProGlnGlnHisLleGlnAspLleThrValValThrAspVal 231
Db      7365 GTATGCTCTACACCCAGCCGAAACACCTGGGTGAGACATATCACCGGTGAGCAGCATGTC 7424
Qy      232 -----AspGlyAspAsnGlyLeuLleAsnTyrGluValGluValAlaAsnGlnThr 249
Db      7425 GCCCAAGACTGTAAACCAACGCGCTGTGACTGGCAGGTG--GTGGCAAT----- 7472
Qy      250 GlyGlnLleGlnLleSerValIleAspGluAspGlyValAlaLeuValAlaValAsnSerGly 269
Db      7473 GTGAGTGTACGCTTGAACCTGCGTATGCGGATCAACAGGTGTTGCACTGACGAACAGGC 7532
Qy      270 AlaGlnGlyThrValThrLleProSerValLysLeuTrpGlnProGlyAlaAlaTyrLeu 289
Db      7533 ACTAGCGGACCTTGGCAAGTGAATCCGCACTTCGCAACCGGCGTGAAGTTATCTC 7592
Qy      290 TyrGlnLeuGlnValAsnLleValGlySerSerGlyAspValValAspThrTyrAsnLeu 309
Db      7593 TATGAACCTGTGCGTC-----ACAGCCAAAAGCCAGACAGATGTATATCTACCCGCTT 7646
Qy      310 AlaThrGlyValArgThrValLysValAlaGlySerGlnPheLeuIleAsnGlyLysPro 329
Db      7647 CGCGTGGCATCTCCGATCAGTGGCAGTGAAGGGCGAACGATTCTTGATTAACCAACAAACCG 7706
Qy      330 PheTyrPheThrGlyPheGlyLysHisGluAspThrAlaValArgGlyLysGlyHisAsp 349
Db      7707 TTCTACTTTACCTGCTTGTGCTGCTATGAAGATCGGCACTGGGTGCAAAAGATTCGAT 7766
Qy      350 ProAlaTyrMetValHisAspPheGlnLeuMetLysTrpLleGlyAlaAsnSerPheArg 369
Db      7767 AACGTCTGATGTGTGACGACCAACGATTAATGATGATGATGATGATGATGATGATGATGAT 7826
Qy      370 ThrSerHisTyrProTyrAlaGluGluValAlaMetAspPheAlaAspArgAsnGlyLleVal 389
Db      7827 ACCCTGCAATTAACCTTAACGCTTAAGAGATGCTGATGCTGGCAATGATGAACATGAGATCGTG 7886
Qy      390 ValIleAspGluThrProAlaValGlyLeuAsnIleAlaLeu--MetGlyValSerGlu 408
Db      7887 GTGATGATGAATACCTGCTGCTGCTGCTTAACTCTCTTAAAGCATTTGGTTTGAAGCG 7946
Qy      409 SerGlyAlaPro--GlnThrPheThrProAspAlaIleAsnAspLysThrGlnGluAla 427
Db      7947 GGCACAAAGCCGAAAGAACTGTACAGCAAGACGCTCAACCGGAAACCTCAGCAAGCG 8006
Qy      428 HisLysGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHisAlaSerValValMet 447

```

Db	8007	CACCTTACGCGGATTTAAAGACCTGATAGCGGGTGTACAAAACCAACCCAGCGTGTATG	8066
Qy	448	TrpSer11ealaaenGluProalSerHisgluAspGlylaargGluTyrrPheGluPro	467
Db	8067	TGGAGTATTCGCAACGAACCGGATACCGTCGCCAAGGTGACGGGAATATTTGCCGCA	8126
Qy	468	LeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGly	487
Db	8127	CTGGCGGAAGCAACGGCTTAACTGACCCCGACGGCTCGGATCACCTGGCTCAATGTATG	8186
Qy	488	ThrIalathTyrrGlnLeuAspArg11SerAspLeuPheAspValSerCysI1easArg	507
Db	8187	TTCTGCGACGCTCACACCGATACCATCGCATCTCTTGTATGTGCTGTGCTGAACCGT	8246
Qy	508	TyrPheGlyTrpTyrrSerGlnThrGlyAspLeuGluGlnIleGlnAlaAlaLeuGlnIlys	527
Db	8247	TATTACGATGCTGATGTCTCCAAACGCGCATTTGGAAACGCAAGAAAGCTACTGGAAAA	8306
Qy	528	GluLeuHisGlyTrpGlnGlnIlysPheHisArgProI1eAlaMetThrGluTyrrGlyAla	547
Db	8307	GNACSTTCGGCGCTGCGAGAGAAACTGCATCAGCGATTAATCATCACCGAATACCGCGCTG	8366
Qy	548	AspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTrpSerGlnGluPheGlnVal	567
Db	8367	GATACGTTAGCCCGGGCTGCACTCAATGTACACCGCAAGTGGAGTGAAGATATCAAGCT	8426
Qy	568	GlnMetLeuAspMetTyrrHisArgValPheAspArg11eGluSerMetAlaGlyGlnHis	587
Db	8427	GCATGGCTGGATATGTATCACCGCGCTTTGATCCGTCACACCGCGTCGTGGTGAACAG	8486
Qy	588	ValTrpAsnPheAlaAspPheGlnThrAsnLeuGlyI1eIleArgValAspGlyAsnIlys	607
Db	8487	GTAATGAAATTCGCCGATTTTGGCACTTCGCAAGGCATATTCGCGGTTGGCGGTACAG	8546
Qy	608	LysGlyValPheThrArgAspArgIysProIysAlaAlaAlaHisSerLeuArgAlaArg	627
Db	8547	AAAGGGAATCTTCACTCCGACCGCAAAACCGAAGTCGCGCGCTTTTCTGCTCAAAAACGC	8606
Qy	628	TrpThrSerI1eAsp	632
Db	8607	TGGATGCGCATGAC	8621


```

Db      8549 TTCTACTTACTGCTTGGTCGTCATGAGATCGGACTGGCTGGCAAGATTCGAT 8490
Qy      350 ProAlATyMeTAlaHsApRheGlnLeuMeTyTrpIlleGlyAlaAsnSerPheArg 369
Db      8489 AACGTCTGATGTGCGACGACCGCATTAATGACATGGAGTGGCGCAACTCCTACCGT 8430
Qy      370 ThSerIATyProTyrAlaGluGluValaMeCAspPheAlaAspArgAsnGlyIleVal 369
Db      8429 ACCTCGCATTAACCTTACGCTGAGAGATGCTCGACTGGCGCATGACATGGCATGTG 8370
Qy      390 ValIleAspGluThrProAlaValaGlyLeuAsnIleAlaLeu--MeCglyValSerGlu 408
Db      8369 GTGATTCATGAACCTGCTGCTGCTTACCTCTTACCTCTTACGATTTGGTTGCAAGCG 8310
Qy      409 SerGlyAlaPro--GlnThrPheThrProAspAlaIleAsnAspGlyTrpGlnGluAla 427
Db      8309 GGCACACAGCCGAGAAAGCTGTACAGCGAAGAGCGACTCAACGGGAAAGTCAAGCAGCG 8250
Qy      428 HsLysGlnAlaIleArgGluLeuIleAlaArgAspIleAsnIleAlaSerValValMet 447
Db      8249 CACTTACAGCGCATTAAGAGCTGATGACCGGTGACAAAACCCAGCGGTGGTATG 8190
Qy      448 TrpSerIleAlaAsnGluProAlaSerHsGluAspGlyAlaArgGluTrpPheGluPro 467
Db      8189 TGGAGTATTCGCAACGAGACCGGATACCCGTCGCAAGTGCACGGAAATTTTCGCCCA 8130
Qy      468 LeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGly 487
Db      8129 CTGCGCGAGCAACGCGCTAACTGACCCGACGCGTCCGATCACTCCGTCACATGTATG 8070
Qy      488 ThrAlaThrTrpGlnLeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArg 507
Db      8069 TTCTGCGACGCTACACCGATACCATGAGGATCTTTATGTGTCTGTCTGTAACCGT 8010
Qy      508 TyPheGlyTrpTySerGlnThrGlyAspLeuGluGluAlaGluAlaIleGluGluLys 527
Db      8009 TATTACGATGATGATGTCGAACGCGGATTTGGAAACGCGACAGAAAGTACTGAAAAA 7950
Qy      528 GlnLeuHsGlyTrpGlnGluLysPheHsArgProIleValMetThrGluTrpGlyAla 547
Db      7949 GAATCTTCTGCGCTGCGAGGAAACTGCATCAGCCGATTCATCAGCCGATTCGCGCTG 7890
Qy      548 AspThrLeuAlaGlyLeuHsSerIleLeuGlyLeuProTrpSerGluGluPheGlnVal 567
Db      7889 GATAACGTAGCCGCGCTGACCTCATATACACGACATGTGTAGTGAAGATACAGTGT 7830
Qy      568 GlnMetLeuAspMetTyHsArgValPheAspArgIleGluSerMetAlaGlyLysHs 587
Db      7829 GCATGCGTGAATGATATCACCGGCTTTGATCGCTCAGCCGCTCGGTGAACAG 7770
Qy      588 ValTrpAsnPheAlaAspPheGlnThrAsnLeuGlyIleIleArgValaAspGlyAsnLys 607
Db      7769 GATTCGAAATTCGCCGATTTTGGACCTCGCAAGGCAATATTGCGCGCTGCGGTAACAG 7710
Qy      608 LysGlyValaPheThrArgAspArgLysProGlyAlaIleAlaHsSerLeuAsnGlyArg 627
Db      7709 AAAGGAGTCTTCTGCGAGCCGAAACGAAAGTGGCGGCTTTTCTGTCGCAAAAACGC 7650
Qy      628 TrpThrSerIleAsp 632
Db      7649 TGGACTGCGCATGAAC 7635

```

```

; FILE REFERENCE: 70149USNP
; CURRENT APPLICATION NUMBER: US/10/680,824A
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/418596
; PRIOR FILING DATE: 2002-07-10
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
; LENGTH: 10011
; TYPE: DNA
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: plasmid pEB10
US-10-680-824A-19

Alignment Scores:
Pred. No.: 1.23e-168 Length: 10011
Score: 1714.00 Matches: 331
Percent Similarity: 70.50% Conservative: 92
Best Local Similarity: 55.17% Mismatches: 161
Query Match: 51.10% Indels: 16
DB: Gaps: 9

US-10-757-093-4 (1-634) x US-10-680-824A-19 (1-10011)
Qy      40 ValArgProGlnArgHsSerArgGluLeuValaAsnLeuAspGlyLeuTrpLysPhe 59
Db      2073 GTCCGCTCTGTAGAACCCCAACCGTGAATGATCAAC-----CGTTGGTGGAAACGGCTTA 2014
Qy      60 AlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnProTrpThrAlaProLeu 76
Db      2013 AGCTGTGATCGCAAAACTGTGATTCATCAG-----CGTTGGTGGAAACGGCTTA 1960
Qy      77 ProLysGlyLeuGluCysProValProAlaSerTyAsnAspIlePheIleSerArgLys 96
Db      1959 CAAGAAAGCCGCGCAATTCGCTGCCAGCGAGCTTTAACATGATCATGCTCCGATGCAAGAT 1900
Qy      97 HsHsAspHsValaGlyTrpValTyTrpGlnArgGluValaIleValProLysGlyTrp 116
Db      1899 ATTGTAATTAATCGGAGCAACGCTGTGTATCAGCGGAAAGTCTTTATACGAAAGGTTGG 1840
Qy      117 SerGlnGluArgTyLeuValaArgAlaGluSerAlaThrHsHsGlyArgIleTyVal 136
Db      1839 GCAGGCGAGGATATCGTGTGCTTCGATCGCGTCACTCATTAACGCAAAAGTGGTTC 1780
Qy      137 AsnAsnArgLeuValaIleGluHsValaGlyTyTrpThrProPheGluAlaAspValThr 156
Db      1779 AATTAATCAGAAAGTATGAGCATCAGGCGGCTATACCGCATTTGAAGCCGATGCACG 1720
Qy      157 GlnLeuValaAlaProGlyGluLysPheArgLeuThrIleGlyValaIleAsnGluLeuThr 176
Db      1719 CCGTATGTTATTCGCGGAAAGTGAATGATACCGTTGTGTGAACAGAACTGAAC 1660
Qy      177 HsGluThrIleProProGlyLysIleThrThyGlyAsnAlaThrGlyLysArgIleGln 196
Db      1659 TGGCAACATATCCCGCGGAAATGTGATTAAC---GACCAAAAACGCAAGAAAGCAG 1603
Qy      197 ThrTyGlnHsAspPheTyAsnTyAlaGlyLeuAlaArgSerIleTrpLeuTySer 216
Db      1602 TCTTACTTCATGATTTCTTAACTATGCGGGATCATCGACGTAATGCTCTACAC 1543
Qy      217 ValProGlnGlnHsIleGlnAspIleThrValaIleThrAspVala-----AspGlyAsp 234
Db      1542 ACGCCAACAACCTGGGTGAGCATATCACCGGTGTACCGATGTCGCGCAAGCTGTAAAC 1483
Qy      235 AsnGlyLeuIleAsnTyArgGluValaIleValaIleAsnGlnThrThyGlyGlnIleGlnIle 254
Db      1482 CAGCGCTGTGTACTGCGAGGTG---GTGGCCAAI-----GCTATGTCAACGCTT 1435
Qy      255 SerValIleAspGluAspGlyAlaIleValaIleValaIleAsnGlyAlaGlnGlyThrVal 274
Db      1434 GAATCGCGATGCGGATCAACAGTGTGTGCAACTGCAAGGCACTAGCGGAGCTTTG 1375

```

```

RESULT 13
US-10-680-824A-19/c
; Sequence 19, Application US/10680824A
; Publication No. US20040133937A1
; GENERAL INFORMATION:
; APPLICANT: Boudreau, Eric
; APPLICANT: Gu, Weining
; APPLICANT: De Fremond, Anic
; APPLICANT: Heifetz, Peter
; TITLE OF INVENTION: Plasmid Transformation

```



```

Db      1393 TGCAGACTATCCCGCGGGAATGATTAAC---ACAGAAACGGCAAGAAACAG 1449
Qy      197  ThrTglnHisAspPheTyrAsnTyrAlaGlyLeuAlaArgSerIleTyrPheTyrSer 216
Db      1450 TCTTACTTCATGATTTCTTAACTAATCGGGAATCCATCGACCGTAATGCTCTTAAC 1509
Qy      217  ValProGlnGlnHisIleGlnAspIleThrValValThrAspVal-----AspGlyAsp 234
Db      1510 ACCCGCAACACCTCGGCGTGAAGATATCACCGTGTGACGATGCGCGCAAGACTGTAAAC 1569
Qy      235  AsnGlyLeuIleAsnTyrGluValGluValAlaAsnGlnThrGlyGlnIleGlnIle 254
Db      1570 CACGCGCTGTGACTGCGCAGGTG---GTGGCCAAAT-----GGTGAATGACGGT 1617
Qy      255  SerValIleAspGluAspGlyAlaIleValAlaValAspGlyAlaGlnGlyThrVal 274
Db      1618 GAACCTGGATCGGATTCACACAGTGTGCACTGACACAGGCAAGCACTAGCGGACTTTG 1677
Qy      275  ThrIleProSerValIleAspTyrGlnProGlyValAlaIleTyrLeuTyrGlnLeuGlnVal 294
Db      1678 CAAGTGTGAATCCGCACCTCTGCAACCGGCGTGAAGTATCTCTATGAACCTGTGCTC 1737
Qy      295  AsnIleValGlySerSerGlyAspValValAspThrTyrAsnLeuAlaThrGlyValArg 314
Db      1738 -----ACAGCCAAAGCCAGACAGAGTGTGATCTACCCGCTGCGTCGCGCATCCG 1791
Qy      315  ThrValIleValAlaGlySerGlnPheLeuIleAsnGlyIleProPheTyrPheThrGly 334
Db      1792 TCAGTGTGACATGAAGGCGCAAGTCTCTATTAACCAACAAACCGTTCTCTTAACGCGC 1851
Qy      335  PheGlyValHisGluAspThrAlaValAlaArgGlyValAspGlyHisAspProAlaTyrFmeVal 354
Db      1852 TTTGGTGTATGAAGATGCGGACTTCTGCGCAAGAAAGATTCAATACTGCTGATGATG 1911
Qy      355  HisAspPheGlnLeuMetLysTyrIleGlyAlaAsnSerPheArgThrSerHisTyrPro 374
Db      1912 CACGACCAACCATTAATGATGATGATGGGCAACTCTTAACGCTACCTGCACTTAACCT 1971
Qy      375  TyrAlaGlnGluValMetLysPheAlaAspArgAsnGlyIleValValIleAspGluThr 394
Db      1972 TAGCGTGAAGATGCTCGACTGGCATGCAATGCAATCGGATGATGATGAAACT 2031
Qy      395  ProAlaValGlyLeuAsnIleAlaLeu---MetGlyValSerGluSerGlyAlaPro--- 412
Db      2032 GCGTGTGTGGCTTTAACTCTCTTGAAGCTGTTTCCGAAGCGGCAACAACCGCAA 2091
Qy      413  GlnThrPheThrProAspAlaIleAsnAspLysThrGlnGluAlaHisLysGlnAlaIle 432
Db      2092 GAACGTGAACAGCAAGGACGATCAACGCGGAAACTCAGCAAGCGCACTTAACGCGCAT 2151
Qy      433  ArgGluLeuIleAlaIleAspLysValAsnHisAlaSerValAlaMetTyrSerIleAlaAsn 452
Db      2152 AAAGAGCTGTATAGCGCTGACAAACACCAAGCTGTGATGTGAGATATGGCCAAC 2211
Qy      453  GluProAlaSerHisGluAspGlyAlaArgGluTyrPheGluProLeuThrAsnLeuThr 472
Db      2212 GAACCGGATACCGCTCGCAAGGTCAGGGAATATTCGCGCCACTGCGCGAAGCAACG 2271
Qy      473  ArgGlnLeuAspProThrArgProIleThrPheHisAsnValGlyThrAlaThrTyrGln 492
Db      2272 CGTAACTGACACCGACGCGTCCGATCACTGCGTCAATGTAATGTTCTTCGACCGCTAC 2331
Qy      493  LeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArgTyrPheGlyTyr 512
Db      2332 ACCGATTAACCTACGCGATCTCTTGAATGCTGCTGCTGAACCGTTTATCGGATGTAT 2391
Qy      513  SerGlnThrGlyAspLeuGluValAlaGluAlaAlaLeuGluLysGluLeuHisGlyTyr 532
Db      2392 GTCCAAAGCGCGATTTGAAACGCGCAAGAGTACTGGAAGAAAGAACTTTCGCGCTGG 2451
Qy      533  GlnGluLysPheHisArgProIleValMetThrGlyTyrGlyAlaAspThrLeuAlaGly 552

```

```

Db      2452 CAGAGAACTGATCAGCCGATTAATCATACCGGAATACGCGCTGTGATACCGCGG 2511
Qy      553  LeuHisSerIleLeuGlyLeuProTyrSerGluGluPheGlnValGlnMetLeuAspMet 572
Db      2512 CTGACTCAATGATACACCGCATGATGAGTGAAGATGATGATGCTGATGCTGATATG 2571
Qy      573  TyrHisArgValPheAspArgIleGluSerMetAlaGlyGluHisValTyrPheAspAla 592
Db      2572 TATCACCGCGCTTTTGAATGCGTCAAGCGCGCTGTGCTGTAACAGGTATGGAATTCGCC 2631
Qy      593  AspPheGlnThrAsnLeuGlyIleIleArgValAspGlyAsnLysGlyValPheThr 612
Db      2632 GATTTTGCAGCTTCGCAAGGCAATATTCGCGGTGCGGTAAACAAAGGATCTTCAC 2691
Qy      613  ArgAspArgLysProLysAlaAlaAlaHisSerLeuArgAlaArgTyrPheSerIleAsp 632
Db      2692 CGGACCGCAACCAAGATGCGCGCTTTTCTGCTGCAAAACGCTGACGATGCAATGAC 2751

RESULT 15
US-10-424-638-1/c
; Sequence 1, Application US/10424638
; Publication No. US20030175245A1
; GENERAL INFORMATION:
; APPLICANT: BROUGH, DOUGLAS E
; APPLICANT: KING, C R
; APPLICANT: KOVESDI, IMRE
; APPLICANT: SCHAIKLE, JASPER J
; TITLE OR INVENTION: REPLICATION DEFICIENT ADENOVIRAL TNF VECTOR
; FILE REFERENCE: 222226
; CURRENT APPLICATION NUMBER: US/10/424, 638
; PRIORITY FILING DATE: 2003-04-25
; PRIOR APPLICATION NUMBER: US 09/604,694
; PRIORITY FILING DATE: 2000-06-27
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 32798
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct
; US-10-424-638-1

Alignment Scores:
Pred. No.: 6,64e-168 Length: 32798
Score: 1714.00 Matches: 331
Percent Similarity: 70.50% Conservative: 92
Best Local Similarity: 55.17% Mismatches: 161
Query Match: 51.10% Indels: 16
Gaps: 9

US-10-757-093-4 (1-634) x US-10-424-638-1 (1-32798)
Qy      40  ValArgProGlnArgThrSerSerArgGluLeuValAsnLeuAspGlyLeuTyrPhe 59
Db      31982 GTCCGCTCTGTAAACCCCAACCGGTGAATCAAAAACTCAACGCGCTGTGGCATTC 31923
Qy      60  AlLeu-----AlaSerGlyLeuAsnAspThrAlaGlnProTyrThrAlaProLeu 76
Db      31922 AGCTCGATGCGCAAAACTGTGGAATGATCAAC-----CGTTGCTGCGAAACGCGCTTA 31869
Qy      77  ProLysGlyLeuGluLysProValProAlaSerTyrAsnAspIlePheIleSerArgLys 96
Db      31868 CAAAGAAACCGCGCAATGCTGTGCGACGCAAGTTTAAACATGATGCTCCGATGCAAT 31809
Qy      97  IleHisAspHisValGlyTyrValTyrTyrGlnArgGluValIleValProLysGlyTyr 116
Db      31808 ATTGGAATTAATGCGGGAACAGCTGTGATCAAGCGGAAGTCTTTATCCGAAGGTTGG 31749
Qy      117  SerGlnGluArgTyrLeuValAlaGluSerAlaThrHisHisGlyArgIleTyrVal 136
Db      31748 GCAGCGCAGCGTATGCTGTGCTTTCCGATGCGGTGATGATGCGGCAAAAGTGTGGTTC 31689

```

QY 137 AsnAspArgLeuValAlaGluHisValGlyIleTyrThr-ProPheGluAlaAspValThr 156
 DB 31688 AATAAATCAGAAAGATGATGAGCATCAGGGCGCTATACCCATTGAAACCCAGATGACAG 31629
 QY 157 GluLeuValAlaAspGlyIleLysPheArgLeuThrIleGlyValAlaAsnGluLeuThr 176
 DB 31628 CCGTATGATATTCGCCGAAAGTACGTATACCCGTTTGTGTGAACAACAGAACGAAAC 31569
 QY 177 HisGluThrIleProProGlyIleThrThrGlyAsnAlaThrGlyLysArgIleGln 196
 DB 31568 TGGCAGACTATCCCGCGGGAATGATGATACC---GACGAAAAACGCAAGAAAAAGCAG 31512
 QY 197 ThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaArgSerIleTyrLeuTyrSer 216
 DB 31511 TCTTACTTCATATTTCTTAACTATGCGGAATCCATGCGAGCGTAAATGCTACACCC 31452
 QY 217 ValProGlnGlnHisIleGlnAspIleThrValIleThrAspVal-----AspGlyAsp 234
 DB 31451 ACCGCCAAGACCTCGGTGAGACGATATCACCGTGTGACGATGTCGCGCAAGCTGTAAAC 31392
 QY 235 AsnGlyLeuIleAsnTyrGluValAlaGluValAlaAsnGlnThrThrGlyGlnIleGlnIle 254
 DB 31391 CACGCGCTGTGTGATGACGAGGTG---GTGGCCAAAT-----GGTGATGTCACGCGT 31344
 QY 255 SerValIleAspGluAspGlyAlaIleValAlaLysAlaSerGlyAlaGlnIleThrVal 274
 DB 31343 GAACTGCGTATCGCATCAACAGGTGTGCACTGGACAAAGACCTAGCGGGAATCTTG 31284
 QY 275 ThrIleProSerValIleLysLeuTyrGlnProGlyValAlaIleTyrLeuTyrGlnLeuGlnVal 294
 DB 31283 CAGTGTGTGAATCCGACCTCTGCGCAACCGGTGAAGTATCTCATGAACTGTGCGTC 31224
 QY 295 AsnIleValIleGlySerSerGlyAspValValAspThrTyrAsnLeuAlaThrGlyValArg 314
 DB 31223 -----ACAGCCAAAGCCAGACAGAGTGTATATCTACCCGCTTCGCGTGGCATCCGG 31170
 QY 315 ThrValIleValAlaGlySerGlnPheLeuIleAsnGlyLysProPheTyrPheThrGly 334
 DB 31169 TCGATGCGAGTGAAGGCGAGCAAGTCTCTGATTAACAACAACCGTTCTTACTTACGGC 31110
 QY 335 PheGlyLysHisGluAspThrAlaValAlaArgGlyLysGlyHisAspProAlaTyrMetVal 354
 DB 31109 TTTGCTGTCATGAAGATGCGGACTTCGTCGCAAGATTCGATTAACGTGATGATGTCG 31050
 QY 355 HisAspPheGlnLeuMetLysTyrIleGlyValAlaAsnSerPheArgThrSerHisTyrPro 374
 DB 31049 CACGACCAACGATTAATGACGTGATGGGCGCAACTCTTACCTGATCTGCAATTACCT 30990
 QY 375 TyrAlaGluGluValIleMetAspPheAlaAspArgAsnGlyIleValAlaIleAspGluThr 394
 DB 30989 TAGCTGAAGATGCTCGACTGGGCAAGTGAACATGCGATCGTGATGATGAACACT 30930
 QY 395 ProAlaValGlyLeuAsnIleAlaLeu--MetGlyValSerGluSerGlyAlaPro--- 412
 DB 30929 GCTGCTGCGGCTTAACTCTCTTAGGCAATGTGTTCCGAGCGGCAACAGCCGAAA 30870
 QY 413 GlnThrPheThrProAspAlaIleAsnAspLysThrGlnGluAlaHisLysGlnAlaIle 432
 DB 30869 GAACTGTACAGCGAAGCGATGCAACGGGAAACTCAAGCAAGCGCACTTACAGGCGATT 30810
 QY 433 ArgGluLeuIleAlaArgAspLysAsnHisAlaSerValIleMetTyrSerIleAlaAsn 452
 DB 30809 AAAAGCTGATAGCGCGTGAACAAAACCAACCAAGCGGTGATGTGAAGTATGGCAAC 30750
 QY 453 GluProAlaSerHisGluAspGlyAlaArgGlyIleTyrPheGluProLeuThrAsnLeuThr 472
 DB 30749 GAAACCGATACCCGTCGCAAGGTGCAAGGGAATATTTCCGCCCACTGGCGGAAGCAACG 30690
 QY 473 ArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGlyThrAlaThrTyrGln 492
 DB 30689 CGTAAACTCGAACCAGCGCGTCCGATCACTGCGCTCAATGTAAATGTTCTGCGACGCTCAC 30630
 QY 493 LeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArgTyrPheGlyTyrPThr 512

DB 30629 ACCGATACCATCAGCGATCTCTTATGTCGTCGCTGAAACCGTTATACGATGATAT 30570
 QY 513 SerGlnThrGlyAspLeuGluGluAlaGluAlaLeuGluLysGluLeuHisGlyTyr 532
 DB 30569 GTCCAAAGCGCGGATTTGGAAACGGCAGAGAGTACTGAAAAAGAACTTGGCCCTGG 30510
 QY 533 GlnGluLysPheHisArgProIleValIleMetThrGluTyrGlyAlaAspThrLeuAlaGly 552
 DB 30509 CAGGAGAAACTGCAATCAGCCGATTTATCATCACCGAAATACCGCGTGAATACGTTACCGGG 30450
 QY 553 LeuHisSerIleLeuGlyLeuProTyrPheSerGluGluPheGlnValGlnMetLeuAspMet 572
 DB 30449 CTGCACTGAATGTACACCGACATGTGAGTGAAGATACAGTGTGATGCTGATATAG 30390
 QY 573 TyrHisArgValPheAspArgIleGluSerMetAlaGlyGluHisValThrAsnPheAla 592
 DB 30389 TATCACCGCGCTTGTGATCGCTCAGCGCGCTGCGGTGAACAGGTATGGAAATTCGCC 30330
 QY 593 AspPheGlnThrAsnLeuGlyIleIleArgValAlaAspGlyAsnLysGlyValPheThr 612
 DB 30329 GATTTTGCACCTCGCAGGCAATATTCGCCGTTGGCGGTAAACAAGAGGATCTTCACT 30270
 QY 613 ArgAspArgLysProLysAlaAlaHisSerLeuArgAlaArgTyrThrSerIleAsp 632
 DB 30269 CCGACCGCAACCGAAGTGGCGGCTTTCTGTGCAAAAAACGCTGATGCGCATGAAC 30210

Search completed: March 22, 2005, 20:39:38
 Job time : 966 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: March 22, 2005, 18:06:15 ; Search time 284 Seconds
(without alignments)
3652.815 Million cell updates/sec

Title: US-10-757-093-4

Perfect score: 3354
Sequence: 1 MFFLGSLSLSLAPSLGTP.....RKPKAAHSLRATWSIDKN 634

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 81813359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODS=frame.p2n.model -DRV=xlp
-Q=/cgnt2_1/USPTO.epool_p/US10757093/runat_18032005_164456_27776/app_query.faeta_1.775
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bites -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=0 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10757093 @CGN_1_1105 @runat_18032005_164456_27776 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA.*
1: /cgnt2_6/prodata/1/ina/5A_COMB.seq.*
2: /cgnt2_6/prodata/1/ina/5B_COMB.seq.*
3: /cgnt2_6/prodata/1/ina/6A_COMB.seq.*
4: /cgnt2_6/prodata/1/ina/6B_COMB.seq.*
5: /cgnt2_6/prodata/1/ina/PCUS_COMB.seq.*
6: /cgnt2_6/prodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1721.5	51.3	3169	3	US-08-630-820-5
2	1721.5	51.3	3169	4	US-09-273-453-5
3	1714.5	51.1	32798	4	US-09-604-694B-1
4	1712.5	51.1	3035	2	US-08-723-624-18
5	1711	51.0	3824	2	US-08-723-624-19
6	1709	51.0	4652	4	US-09-893-525-36
7	1706	50.9	5390	4	US-09-893-525-41
8	1706	50.9	5418	4	US-09-893-525-38
9	1703	50.8	11978	4	US-09-792-568-8
10	1703	50.8	12438	4	US-09-792-568-9
11	1692	50.4	8012	3	US-09-182-117-1
12	1692	50.4	8012	4	US-09-434-039A-1

13	1692	50.4	8418	3	US-09-182-117-5	Sequence 5, Appl1
14	1692	50.4	8418	4	US-09-434-039A-5	Sequence 5, Appl1
15	1692	50.4	8798	3	US-09-182-117-4	Sequence 4, Appl1
16	1692	50.4	8798	4	US-09-434-039A-4	Sequence 4, Appl1
17	1673.5	49.9	2141	4	US-09-445-283C-27	Sequence 27, Appl1
18	1673.5	49.9	8076	4	US-09-532-806-2	Sequence 2, Appl1
19	1667.5	49.7	4947	4	US-09-118-276-21	Sequence 21, Appl1
20	1667.5	49.7	5642	1	US-08-118-772A-2	Sequence 2, Appl1
21	1667.5	49.7	5897	4	US-09-097-319A-26	Sequence 26, Appl1
22	1667.5	49.7	5897	4	US-09-643-971A-26	Sequence 26, Appl1
23	1667.5	49.7	6898	3	US-09-097-319A-27	Sequence 27, Appl1
24	1667.5	49.7	6898	4	US-09-643-971A-27	Sequence 27, Appl1
25	1667.5	49.7	9002	4	US-09-532-806-3	Sequence 3, Appl1
26	1667.5	49.7	9299	3	US-09-643-971A-15	Sequence 15, Appl1
27	1667.5	49.7	9299	4	US-09-643-971A-15	Sequence 15, Appl1
28	1667.5	49.7	9335	3	US-09-097-319A-19	Sequence 19, Appl1
29	1667.5	49.7	9335	4	US-09-643-971A-19	Sequence 19, Appl1
30	1667.5	49.7	9408	4	US-09-643-971A-16	Sequence 16, Appl1
31	1667.5	49.7	9408	3	US-09-643-971A-16	Sequence 16, Appl1
32	1667.5	49.7	10160	4	US-09-097-319A-8	Sequence 8, Appl1
33	1667.5	49.7	10160	4	US-09-643-971A-8	Sequence 8, Appl1
34	1667.5	49.7	11784	4	US-09-097-319A-9	Sequence 9, Appl1
35	1667.5	49.7	11784	4	US-09-643-971A-9	Sequence 9, Appl1
36	1667.5	49.7	11991	3	US-09-097-319A-10	Sequence 10, Appl1
37	1667.5	49.7	11991	4	US-09-643-971A-10	Sequence 10, Appl1
38	1666	49.7	7742	2	US-08-882-704A-4	Sequence 4, Appl1
39	1666	49.7	7742	3	US-09-151-957-4	Sequence 4, Appl1
40	1664.5	49.6	4544	4	US-09-488-270A-1	Sequence 1, Appl1
41	1659	49.5	2438	6	5432081-1	Sequence 1, Appl1
42	1659	49.5	2438	6	5432081-1	Sequence 1, Appl1
43	1648.5	49.2	4284	1	US-08-525-507-14	Sequence 14, Appl1
44	1631.5	48.6	2633	1	US-08-452-267-2	Sequence 2, Appl1
45	1631.5	48.6	2633	3	US-09-123-644-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1
US-08-630-820-5
Sequence 5, Application US/08630820
Patent No. 6008023
GENERAL INFORMATION:
APPLICANT: OPPER, Martin
APPLICANT: BOSSLET, Klaus
APPLICANT: CZECH, Joerg
TITLE OF INVENTION: CYTOPLASMIC EXPRESSION OF ANTIBODIES,
TITLE OF INVENTION: ANTIBODY FRAGMENTS AND ANTIBODY FRAGMENT FUSION MOLECULES
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSER: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,820
FILING DATE: 10-APR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE 19513676.4
FILING DATE: 11-APR-1995
ATTORNEY/AGENT INFORMATION:
NAME: GRANADOS, Patricia D.
REGISTRATION NUMBER: 33,683
REFERENCE/DOCKET NUMBER: 18748/306
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300

TELEFAX: (202) 672-5399
 TELEX: 904136
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3169 base pairs
 type: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: circular
 MOLECULE TYPE: cDNA
 HYPOTHEetical: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Enterobacteriaceae: Escherichia coli
 STRAIN: PR4210
 IMMEDIATE SOURCE:
 CLONE: pTc99 dicistr. Fab/B.c.-Beta-Gluc
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 3..641
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 666..3162
 OS-08-630-820-5

Alignment Scores:

Pred. No.:	3.02e-175	Length:	3169
Score:	1721.50	Matches:	339
Percent Similarity:	68.17%	Conservative:	98
Best Local Similarity:	58.89%	Mismatches:	179
Query Match:	51.33%	Indels:	25
DB:	3	Gaps:	11

US-10-757-093-4 (1-634) X US-08-630-820-5 (1-3169)

Qy	8	SerLeuLeuSerLeuLeuAlaAlaProSerLeuGlyThrProAla-----AlaArg	23
Db	1233	AGCGTGGAGACCGTGCCTCCAGCAAGCTTGGGACCCAGACCTACACCTGCAGCTGAAT	1292
Qy	24	HisPheProArgAsnGluMetThrGlnHisGluGlnProLeu-----Ile	38
Db	1293	CACAAGCCACCAACACCAAGGTGGACAGAAGAGTTCAGCTCGGTTCGTTCGGTTCC	1352
Qy	39	LysValAlaArgProGlnArgThrSerSerArgGluLeuValAsnLeuAspGlyLeuTrpLys	58
Db	1353	ATGGTACGCTCGTATAGAAACCCCAACCCGTGAATCAAAAACTGCAGCGGCTGTGGCA	1412
Qy	59	PheAlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnProTrpThrAlaPro	75
Db	1413	TTCAGTCTGATCCGCAAACTGTGGATATATACG-----CGTTGGTGGAAAGCCG	1466
Qy	76	LeuProLysGlyLeuGluCysProValProAlaSerTrpAsnAspIlePheIleSerArg	95
Db	1467	TTACAGAAACCCGGCGCAATTGCTGTCCAGGCAAGTTTAAAGATCAAGTCCGCAGTCA	1526
Qy	96	GluIleHisAspHisValGlyTrpValTrpTrpGlnAsnGluValIleValProLysGly	115
Db	1527	GATATTCTGAATTATATGCGGGCAAGCTCGTATACGCCGGAAGCTTATATCCGAAAGT	1586
Qy	116	TrpSerGlnGluArgTrpLeuValArgAlaGlnSerAlaThrHisGlyValArgIleTrp	135
Db	1587	TGGGAGGCCACCGCATGTGTGGTGGTTCGATGGCGTCACTCATTTAGCGCAAGTGGG	1646
Qy	136	ValAsnAsnArgLeuValAlaGluHisValGlyGlyTrpThrProPheGluValAspVal	155
Db	1647	GTCATATATACGAAGATGATGGAGCATCAGGCGCGCTATAGCCATTTGAAGCGCATTC	1706
Qy	156	ThrGluLeuValAlaProGlyGlyLysPheArgLeuThrIleGlyValAsnAsnGluLeu	175
Db	1707	ACGCGATGATTATATGCGGGGAAAAGTACGATACACCGTTTGGTGAACAACAACACTG	1766
Qy	176	ThrHisGluThrIleProProGlyLysIleThrThrGlyAsnAlaThrGlyLysArgIle	195
Db	1767	AACCTGCAGCATATCCCGCCGGGAATGGTATTACC--GACGAAAACGGCAAGAAAG	1822

QY	196	GLNTHRGLNHSASPPEYRANSTYVALAGLYLEUAAASERILETPLEUTYR	215
DB	1824	CAGCTTACTTCCTAATATTTCTTTAACTATGCCGGATCCATGCCAGCGTAATGCTTAC	1883
QY	216	SErVALPRGGLNGLNHSILEGINASPIELTHrVALThrAspVAL-----AspGLY	233
DB	1884	ACCACCGCGAACAACCTGGGTGCAGATATACCGGTGTGCAGCATGTCCGGCAAGACTGT	1943
QY	234	ASpASnGLYleuILEaenTYrGLVVALaLEaSnGLNThrThrGLNILEGLN	253
DB	1944	AACCAACCGCTGTGTGACTGCACAGT---GTGGCCAAT-----GGTATGTACG	1991
QY	254	ILESerVALILEASpGLUASpGLVALaILEVALaLEuVALaSERGLYalAGNGLYThr	273
DB	1992	GTTGAACGTGGGTATGCCGATCAACAGGTGTTCACATCGACACAAAGCACTAAGCGGAGCT	2051
QY	274	VALThrILEProSerVALYLSLEUTPRGLNPRGGLYalALaTYrLEuTYrGLNILEGLN	293
DB	2052	TTGGAAATGGTGAATCCGACCTTGCAACCGGGTAAGTTATCTCTAAGACTGTGC	2111
QY	294	VALaSnILEVALGLYSErSERGLYASpVALaLEAspThrTYrASnLEuVALaThrGLYVAL	313
DB	2112	GTC-----ACAGCGAACAACCCAGACAGAGTGAATCTACCCGCTTCGGGTGGCATC	2165
QY	314	ArgThrVALYLSrVALaLAGLYSErGLNPhLEuILEaSnGLYLSrProPheTYrPheThr	333
DB	2166	CGGTCACTGGCAGTGAAGGGCGAACAAGTCTCTGATTAAACAACAAACCGTTCTACTTAACT	2225
QY	334	GLYPheGLYLSHISGLIAspThrVALaVALArgGLYLSrGLYHISAspProaLetyMet	353
DB	2226	GGCTTGTGTGTCTGAAGAATGCCGACTTACGTGGCAAGATTCCGATACGTCTGATG	2285
QY	354	VALHISAspPheGLNLeuMetLYSTPILGLYAlaAsnSerPheArgThrSerHISTYR	373
DB	2286	GTGCACGACCAACCGATTATATGAAGTGGTGGGCCAACTCTCCCTACTCCGACTTAC	2345
QY	374	ProTYrTALaGLIuGLIuValMeLAspPheVALaAspArgaSnGLYILEVALaILEASpGLU	393
DB	2346	CTTTACGCTGAAGAAGATGCTGACTGGCGAGATGAACATGGCAATCGTGGATATGATGA	2405
QY	394	ThrProaLValGLYLeuASnILEaLeu---MetGLYALSerGLUSeRGLYALaPro	412
DB	2406	ACTGCTGCTGTGGCTTTAACTCTCTTTAGCAATGTGTTCGAAGGGGCAACAAGCCG	2465
QY	413	---GLNThrPheThrProaSPaLILEaSnAspLYSThrGLNGLIuALHISGLYGLIuAL	431
DB	2466	AAAGAACTGTACACGCGAAGGCGAGTCAACCGGGAACCTCAGCAAGCGCAATTACAGCG	2525
QY	432	ILEArgGLIuLeuILEaArgAspLYASnHISaLESerVALaILEMetTSPSerILEaL	451
DB	2526	ATTAAAGTCTGATAGCGCGTGAACAAAACCAACCAAGCGTGTGATGTGGAGATTTGCC	2585
QY	452	ASnGLUProaLSErHISGLIuASpGLYALaArgGLIuTYrPheGLUProLeuThrASnLeu	471
DB	2586	AACGAACCGGATACCCGTCCGCAAGGTGCACAGGAAATTTTCGGCCACTGGCGGAACA	2645
QY	472	ThrTYrGLIuLeuAspProThrTYrArgProILEThrPheHISaSnVALaGLYThrALaThrTYR	491
DB	2646	ACGGGTAACTCGACCCGACCGCGTCCATCACTCGGCTCAATGTAAATGTTTGGAGAGCT	2705
QY	492	GLIuLeuAspArgILEASpLEuPheAspVALaSerCYuSILEaSnArgTYrPheGLYTrp	511
DB	2706	CACACGATACCATCAAGAGCTCTTTGATGTGTGTGTCCTGAACCGTTATATGAGATGG	2765
QY	512	TYrSerGLNThrGLYAspLEuGLIuALaGLIuALaLLeuGLIuLYrGLIuLeuHISGLY	531
DB	2766	TATGTCCAAAGCGCGCATTTGGAAACGCGAGAGATGACTGGAAAAAGAACTTCTGGCC	2825
QY	532	ThrGLIuGLYLSrPheHISArgProILEValMetThrGLIuTYrGLYALaAspThrLEuVAL	551
DB	2826	TGGCAGAGAAATCTCAACGCGATTTATCTACCGAATATCGGCGTGGATACGTTAGCC	2886

QY 552 GlyLeuHisSerIleLeuGlyLeuProTrpSerGluGlnPheGlnValGlnMetLeuAsp 571
 Db 2886 GGGCTGACATCATATGTAACCGACATGTGAGTGAAGATATCATGTGATGCTGAT 2945
 QY 572 MetTyrHisArgValPheAspArgIleGluSerMetAlaGlyGlnHisValTTPAspPhe 591
 Db 2946 ATGTATCACCGCGCTTTGATTCGTCAGCGCCGCTCGGTGAACAGTATGGAATTTTC 3005
 QY 592 AlaAspPheGlnThrLeuGlyIleIleArgValAspIleValGlyValPhe 611
 Db 3006 GCGGATTTTGCAGCTCGCAAGCATATTTGCGGTCGTAACGAAGGAGATCTTC 3065
 QY 612 ThrArgAspArgLysProLysAlaAlaHisSerLeuArgAlaArgTTPThrSerIle 631
 Db 3066 ACTCGGACCGCAACCGAAGTGGCGGCTTTCTGCTGCAAAAACCGCTGAGCTGGCATG 3125
 QY 632 Asp 632
 Db 3126 AAC 3128
 RESULT 2
 US-09-273-453-5
 ; Sequence 5, Application US/09273453
 ; Patent No. 6602688
 ; GENERAL INFORMATION:
 ; APPLICANT: OPPER, Martin
 ; BOSSLER, Klaus
 ; CZECH, Joerg
 ; TITLE OF INVENTION: CYTOPLASMIC EXPRESSION OF ANTIBODIES,
 ; ANTIBODY FRAGMENTS AND ANTIBODY FRAGMENT FUSION MOLECULES
 ; IN E. COLI
 ; NUMBER OF SEQUENCES: 7
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Foley & Lardner
 ; STREET: 3000 K Street, N.W., Suite 500
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20007-5109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/273,453
 ; FILING DATE: 22-Mar-1999
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/630,820
 ; FILING DATE: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: GRANADOS, Patricia D.
 ; REGISTRATION NUMBER: 33,683
 ; REFERENCE/DOCKET NUMBER: 18748/306
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 672-5300
 ; TELEFAX: (202) 672-5399
 ; TELEX: 904136
 ; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 3169 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: circular
 ; MOLECULE TYPE: cDNA
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Enterobacteriaceae: Escherichia coli
 ; STRAIN: PR4210
 ; IMMEDIATE SOURCE:
 ; CLONE: pTrc99 dicistr. Fab/E.c.-Beta-Gluc
 ; FEATURE:

NAME/KEY: CDS
 LOCATION: 3..641
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 666..3162
 SEQUENCE DESCRIPTION: SEQ ID NO: 5:
 US-09-273-453-5
 Alignment Scores:
 Pred. No.: 3.02e-175 Length: 3169
 Score: 1721.50 Matches: 339
 Percent Similarity: 68.17% Conservative: 98
 Best Local Similarity: 52.89% Mismatches: 179
 Query Match: 51.33% Indels: 25
 DB: 4 Gaps: 11
 US-10-757-093-4 (1-634) x US-09-273-453-5 (1-3169)
 QY 8 SerLeuLeuSerLeuAlaAlaProSerLeuGlyThrProAla-----AlaArg 23
 Db 1233 AGCGTGATACCGTCGCCCTTCACAGCAGCTTGGCAACCAACCTGCAACGCGAAT 1292
 QY 24 HisPheProArgLengLumeThrGlnHisGluGlnProLeu-----Ile 38
 Db 1293 CACAAACCCGACCAACCAAGTGACAAAGAGAGTGAAGCTGCTGCTGCTGCTTC 1352
 QY 39 LysValArgProGlnArgThrSerArgGluLeuValAsnLeuAspGlyLeuTPlvs 58
 Db 1353 ATGGTACGTCCTGTAGAAACCCCAACCCGGAATCAAAAACCTGACGCGCTGGGCA 1412
 QY 59 PheAlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnProTrpTrAlaPro 75
 Db 1413 TTACAGTCTGATGCGCAAAACCTGGAATTTGATGAC-----CGTTGCTGGAAAGCGG 1466
 QY 76 LeuProLysGlyLeuGluCysProValProAlaSerTyrAsnAspIlePheIleSerArg 95
 Db 1467 TTACAGAAAGCCGGGCAATGCTGTCCAGGAGATTTAACATGATCAGTTCGCGATGCA 1526
 QY 96 GluIleHisAspHisValGlyTTPValIYTYrGlnArgGluValIleValProLysGly 115
 Db 1527 GATATTGCAATTAATTCGCGCAACGCTGTGTACACGCGGAACTTTTATACGAAAGT 1586
 QY 116 TrpSerGlnGluArgYrLeuValAlaGlnIleValIleGlyTyrThrProPheGlnAlaAspVal 135
 Db 1587 TGGGACGCGCAGGCGTATTCGTCGCTTCGATGCGGCTCATATTAACGCAAGTGTG 1646
 QY 136 ValAsnAspArgLeuValAlaGlnHisValGlyTyrThrProPheGlnAlaAspVal 155
 Db 1647 GTCAATTAATCAGAAAGTGAAGATGAGCATCAGGGCGGTATTAACCATTTGAAGCGATGTC 1706
 QY 156 ThrGluLeuValAlaProGlyGluLysPheAspGluThrIleGlyValAsnAspGluLeu 175
 Db 1707 ACCCGCTATGTTATTCGCGGAAAGTGAATGATCAGCTTTGTGTGAACAAGATCTG 1766
 QY 176 ThrHisGluThrIleProProGlyLysIleThrGlyAsnAlaThrGlyLysArgIle 195
 Db 1767 AACTGGACAGCTATCCCGCGGGAATGATGATTAAC---GACCAAAACGCAAGAAAG 1823
 QY 196 GlnThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaAspSerIleTPlvs 215
 Db 1824 CAGCTTACTTCATTAATTTCTTAATTCATGCGGGAATCGGATCGACGGAATGCTTAC 1883
 QY 216 SerValProGlnGlnHisIleGlnAspIleThrValIleThrAspVal-----AspGly 233
 Db 1884 ACCACGCGCAACACCTGCTGTGACGATACACCGTGTGACGATGCTCGCGCAAGCTGT 1943
 QY 234 AspAsnGlyLeuIleAsnTyrGluValAlaAsnGlnThrThrGlyGlnIleGln 253
 Db 1944 AACCAACGCTGCTGTACTGCGAGGTG---GTGGCAAT-----GATGATGACAG 1991
 QY 254 IleSerValIleAspGluAspGlyAlaIleValAlaLysAlaSerGlyAlaGlnGlyThr 273
 Db 1992 GTTGAACGTGCGATGCGGATCAACAGTGTGGCAATGGAACAGCAACGCACTACGCGGACT 2051

QY 274 ValThrIleProSerValIleuThrProGluValAlaTyrLeuTyrGluLeuGln 293
 Db 2052 TTGCAAGTGTGAATCCGCACTCTGGCAACCGGGTGAAGGTATCTCTATGAAGCTGGC 2111
 QY 294 ValAsnIleValGlySerSerGlyAspValValAspThrTyrAsnIleAlaThrGlyVal 313
 Db 2112 GTC-----ACAGCCAAAGCCAGACAGGTGTATCTACCCGCTTCGGGTGGCATC 2165
 QY 314 ArgThrValIysValAlaGlySerGlnPheLeuIleAsnGlyLysProPheTyrPheThr 333
 Db 2166 CGGTAGTGGCAGTGAAGGGCAACGTTCTGTATTACCAACAACCGTTCTACTTACT 2225
 QY 334 GlyPheGlyLysHisGluAspThrAlaValArgGlyLysGlyHisAspProAlaTyrMet 353
 Db 2226 GGCCTTGGTCGTCAATGAAGATCGCGACTTACGTGGCAAGATTCGATACGTCTGATG 2285
 QY 354 ValHisAspPheGlnLeuMetLysTrpIleGlyAlaAsnSerPheArgTyrSerHisTyr 373
 Db 2286 GTGACGACGACGACGATTAATGAGTGGATGGGGCCAACTCTACCGTACCTGGCATTC 2345
 QY 374 ProTyrAlaGluGluValMetAspPheAlaAspArgAsnGlyIleValValIleAspGlu 393
 Db 2346 CTTACGCTGAAGATGCTCTGACTGGCAGATGAACATGCGATCGTGGATGTATGA 2405
 QY 394 ThrProAlaValGlyLeuAsnIleAlaLeu---MetGlyValSerGluSerGlyAlaPro 412
 Db 2406 ACTGCTGCTGCGGCTTAACTCTCTTGGCATTGGTTTCGAAGGGGCAACAACCGC 2465
 QY 413 ---GlnThrPheThrProAspAlaIleAsnAspLysThrGlnGluAlaHisLysGlnAla 431
 Db 2466 AAAAGACGTACAGCAAGAAAGCGACGACGCGGAAACTGACGAAGCGCACTTACAGCGCG 2525
 QY 432 IleArgGluLeuIleAlaArgAspLysAsnHisAlaSerValValMetTrpSerIleAla 451
 Db 2526 ATTAAGAGCTGATAGCGCGTGAACAACCAACCCAGCTGTGATGTGAATATTTGCC 2585
 QY 452 AsnGluProAlaSerHisGluAspGlyAlaArgGluTyrPheGluProLeuThrAsnLeu 471
 Db 2586 AACCAACCGGATACCGCTCGCAAGGTGACGCGAAATATTCGCGCACATGCGGAGCA 2645
 QY 472 ThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGlyThrAlaThrTyr 491
 Db 2646 ACGGTAACTCAACCGGACGCGCTCGCATCACTGGCTCAATTAATGTTTCGACGCT 2705
 QY 492 GlnLeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArgTyrPheGlyTyr 511
 Db 2706 CACACCGATCCCATCAGCGATCTTGTATGTCTGTGCTGTAACCGTTATTAACGATGG 2765
 QY 512 TyrSerGlnThrGlyAspLeuGluGluAlaGluAlaLeuGluLysGluLeuHisGly 531
 Db 2766 TATGTCGAAGCGCGGATTTGGAAACGCGCAGAGAGTACTGAAAAAACAACCTTGGCC 2825
 QY 532 TrpGlnGluLysPheHisAspProIleValMetThrGlyTyrGlyValIleAspThrIleuAla 551
 Db 2826 TGGCAGAGAAATGATCATCGCATTCATCAACGGAATACGGCGTGGATACGTTAGCC 2885
 QY 552 GlyLeuHisSerIleLeuGlyLeuProTrpSerGluGlnPheGlnValGlnMetLeuAsp 571
 Db 2886 GGGCTCACTCAATGACACCGCATGTGAGTGAAGATATCAGTGCATGCGTGGAT 2945
 QY 572 MetTyrHisArgValPheAspArgIleGluSerMetAlaGlyGlnHisValIleTrpAsnPhe 591
 Db 2946 ATGTATACCGCGCTCTTATCGCGTCAACCGCGTGTGGTGAACAGGTATAGGAATTC 3005
 QY 592 AlaAspPheGlnThrAsnLeuGlyIleIleArgValAspGlyAsnLysLysGlyValPhe 611
 Db 3006 GCGGATTTTGCACCTTCGCAAGGCAATATTCGCGTGGCGGTAAACAAGAAAGATCTTC 3065
 QY 612 ThrArgAspArgLysProLysAlaAlaHisSerLeuArgAlaArgTrpThrSerIle 631
 Db 3066 ACTCGGACCGCAAAACCGAAGTGGGGCTTTTCTGTGCAAAAACGCTGACGTGGCATG 3125

QY 632 Asp 632
 Db 3126 AAC 3128
 RESULT 3
 US-09-604-694B-1/C
 ; Sequence 1, Application US/09604694B
 ; Patent No. 6579522
 ; GENERAL INFORMATION:
 ; APPLICANT: BROUGH, DOUGLAS E
 ; APPLICANT: KING, C R
 ; APPLICANT: KOVESDI, IMRE
 ; APPLICANT: SCHAIBLE, JASPER J
 ; TITLE OF INVENTION: REPLICATION DEFICIENT ADENOVIRAL TNF VECTOR
 ; FILE REFERENCE: 202028
 ; CURRENT APPLICATION NUMBER: US/09/604,694B
 ; CURRENT FILING DATE: 2003-01-10
 ; NUMBER OF SEQ ID NOS: 2
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 1
 ; LENGTH: 32798
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Synthetic construct
 US-09-604-694B-1
 Alignment Scores:
 Pred. No.: 8,196-173
 Score: 1714.00
 Percent Similarity: 70.50%
 Best Local Similarity: 55.17%
 Query Match: 51.10%
 DB: 4
 Gaps: 9
 US-10-757-093-4 (1-634) x US-09-604-694B-1 (1-32798)
 QY 40 ValArgProGlnArgThrSerSerArgGluLeuValAlaAsnLeuAspGlyLeuTrpLysPhe 59
 Db 31982 GTCCGCTCTGTGAAGAACCCCAACCGGTGAATCAAAAACTGACGCGCTGTGGCATTC 31923
 QY 60 AlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnProTrpThrAlaProLeu 76
 Db 31922 AGTCGTGATCGCAAAACTGTGGAATGATCAG-----CGTGTGTGGAAAGCGCGTTA 31869
 QY 77 ProLysGlyLeuGluCysProValProAlaSerTyrAsnAspIlePheIleSerArgGlu 96
 Db 31868 CAAAGAACCCGGCAATGCTGTGCGCAGGCAAGTTTAAACGATCAGTTCGCCGATGCAGAT 31809
 QY 97 IleHisAspHisValGlyTrpValTyrTyrGlnArgGluValIleValProLysGlyTyr 116
 Db 31808 ATTGTGATATTATGGGGGCAAGCTGTGATACGCGCAAGTCTTTATACGAAGGTTGG 31749
 QY 117 SerGlnGluArgTyrLeuValArgAlaGlnSerAlaThrHisHisGlyArgIleTyrVal 136
 Db 31748 GCAAGCCAGCCTATCGCGCTGCGCTTTCGATCGCGTCACTCATTAACGCAAAAGTGGGTC 31689
 QY 137 AsnAspArgLeuValAlaGlnHisValGlyTyrThrPheProPheGluAlaAspValThr 156
 Db 31688 AATAATACGAAGATGAGAGCATCAGAGGCGGCTATACGCAATTTGAAGCGCATGTCACG 31629
 QY 157 GluLeuValAlaProGlyGluLysPheArgLeuThrIleGlyValAlaAsnGlnLeuThr 176
 Db 31628 CCGTATGTATTATGGCGGAAAGTGAATCATCCGTTGTGTGAACAACGAAGTGAAC 31569
 QY 177 HisGluThrIleProProGlyLysIleThrThrGlyAsnAlaThrGlyLysArgIleGln 196
 Db 31568 TGGCAGACTATCCCGCGGGAAGTGAATACC---GACGAAACGCGCAAGAAAGGCGCG 31512
 QY 197 ThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaArgSerIleTrpLeuTyrSer 216
 Db 31511 TCTTACTTCATGATTTCTTAACTATGCGGGAATCATCGCAGCTGATGCTCTACACC 31452

QY 217 ValProGlnGlnHisIleGlnAspIleThrValValThrAspVal-----AspGlyAsp 234
 DB 31451 AGCCGGAACACCTGGGTGGACGATATACCGTGTGACCGATCTCCGCAAGCTCTGAAC 31392
 QY 235 AsnGlyLeuIleAsnTyrgluValGluValAlaAsnGlnThrGlyGlnIleGlnIle 254
 DB 31391 CACGCGCTGTGTGACTGGCAGGTG---GTGGCCAAAT-----GGTAGTGCACCGGT 31344
 QY 255 SerValIleAspGluAspGlyAlaIleValAlaValAlaSerGlyValGlnGlyThrVal 274
 DB 31343 GAACCTCGATGATCGGATCAACAGGTGTGCACTGACGACAGACGCACTAGCGGACCTTGG 31284
 QY 275 ThrIleProSerValIleuSerTrpGlnProGlyValAlaAlaTyrgluLeuGlnVal 294
 DB 31283 CAAGTGTGAATCCGACCTCTGGCAACCGGTGAAGTTATCTTATGACCTGTGCTC 31224
 QY 295 AsnIleValGlySerSerGlyAspValValAspThrTyrgluLeuAlaThrGlyValArg 314
 DB 31223 -----ACAGCCAAAGCCAGACAGAGTGTGATATCTACCCGCTTGGCGGTGACCTCCG 31170
 QY 315 ThrValValValAlaGlySerGlnPheLeuIleAsnGlyValAspProPheThrPheThrGly 334
 DB 31169 TCAGTGGCAGTGAAGGCGCAACGTTCTGATTACCAACACCGTTTACTTACTGCG 31110
 QY 335 PheGlyValHisGluAspThrAlaValArgGlyValGlyHisAspProAlaTyrgluVal 354
 DB 31109 TTTGGTGTCTATGAATCGGACCTTGGCAAAAGATTGATTAACGTCGATGCTG 31050
 QY 355 HisAspPheGlnLeuMetLeuTyrglyValAlaAsnSerPheArgThrSerHisTyrglyPro 374
 DB 31049 CACGACCAAGCATTAATGATGATGATGGGGCCAACTCTTACCGTACCTGCTACCT 30990
 QY 375 TyrAlaGluGluValMetAspPheAlaAspArgAsnGlyIleValValIleAspGluThr 394
 DB 30989 TACGCTGAAGAGATGCTCGATCGGCAAGATGAACTGCTGTGATGTGAAACT 30930
 QY 395 ProAlaValGlyLeuAsnIleAlaLeu---MetGlyValSerGluSerGlyAlaPro--- 412
 DB 30929 GCTGCTGTGGCTTAACTCTTAAAGCATGCTTTCGAACGGGCAACAGCCGAAA 30870
 QY 413 GlnThrPheThrProAspAlaIleAsnAspLeuThrGlnGluAlaHisGlyGlnAlaIle 432
 DB 30869 GAACCTGACGCGAAGAGGACGATCAACGGGAAACTCAGCAACGCGACTTACAGCGCAT 30810
 QY 433 ArgGluLeuIleAlaArgAspLeuAsnHisAlaSerValValMetTrpSerIleAlaAsn 452
 DB 30809 AAAGACTGATAGCGGTGTGCAAAAACCAACCAAGCGTGTGATGTGATATTCGCAAC 30750
 QY 453 GluProAlaSerHisGluAspGlyAlaArgGluTyrgluPheGluProLeuThrAsnLeuThr 472
 DB 30749 GAACCGGATACCGCTCGGCAAGGTGCAAGGAAATATTCGCGCACTGGCGGAAGCAACG 30690
 QY 473 ArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGlyThrAlaThrTyrgln 492
 DB 30689 CGTAACTCGACCGACCGACGCTCGATCAGCTGCTGATATGTTTCGCAACGCTCAC 30630
 QY 493 LeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArgTyrgluTyrgly 512
 DB 30629 ACCGATACCATCAGCATCTCTTGTGATGTGCTGTGCTGCAACCGCTTATTCGATGTAT 30570
 QY 513 SerGlnThrGlyAspLeuGluGluValGluAlaIleGlnGluGluLeuHisGlyTyrgly 532
 DB 30569 GTCCAAAGCGGCGATTGGAAACGCGCAAGAAAGTACTGAAAAAGAACTTTCGCGCTGG 30510
 QY 533 GlnGluLeuPheHisArgProIleValMetThrGlyTyrglyAlaAspThrLeuAlaGly 552
 DB 30509 CAGGAAGAACTGATCAGCGGATTAATCATCAGCAATACCGCGCTGATACGTCGCGG 30450
 QY 553 LeuHisSerIleLeuGlyLeuProTrpSerGluGluPheGlnValGlnMetLeuAspMet 572
 DB 30449 CTGCACTCATATGACCAACGATGTGAGTGAAGATGAAAGATGTCATGCTGCGATATG 30390
 QY 573 TyrHisArgValPheAspArgIleGluSerMetAlaGlyGluHisValTrpAsnPheAla 592

DB 30389 TATCACCGGCTGTTGATGCGTCAGCGCGCTGCTGGAACAGGATGGAATTTCGCC 30330
 QY 593 AspPheGlnThrAsnLeuGlyIleIleArgValAspGlyAsnTyrglyValPheThr 612
 DB 30329 GATTTCGCACTCTCGCAAGCATTTGGCGGTGTGGCGGTAACAGAAAGGATCTTCACT 30270
 QY 613 ArgAspArgLeuProGlyAlaAlaHisSerLeuArgAlaArgTrpThrSerIleAsp 632
 DB 30269 CGGACCGCAACCGAAGTCGGCGCTTTCGCTGCAAAAACGCTGACATGCAAGAC 30210
 RESULT 4
 US-08-723-624-18
 ; Sequence 18, Application US/08723624
 ; Patent No. 5861277
 ; GENERAL INFORMATION:
 ; APPLICANT: Rose, Alan B.
 ; APPLICANT: Laet, Robert L.
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ENHANCING
 ; TITLE OF INVENTION: THE EXPRESSION OF GENES IN PLANTS
 ; NUMBER OF SEQUENCES: 21
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: ARNOLD, WHITE & DURKEE
 ; STREET: P.O. Box 4433
 ; CITY: Houston
 ; STATE: TX
 ; COUNTRY: USA
 ; ZIP: 77210
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/723,624
 ; FILING DATE: Concurrently Herewith
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Parker, David L.
 ; REGISTRATION NUMBER: 32,165
 ; REFERENCE/DOCKET NUMBER: BTIP:002
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 512/418-3000
 ; TELEFAX: 512/474-7577
 ; INFORMATION FOR SEQ. ID NO: 18:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 3035 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-723-624-18
 Alignment Scores:
 Pred. No.: 2,64e-174 Length: 3035
 Score: 1712.50 Matches: 333
 Percent Similarity: 69.23% Conservative: 99
 Best Local Similarity: 53.37% Mismatches: 169
 Query Match: 51.06% Indels: 23
 DB: 2 Gaps: 10
 US-10-757-093-4 (1-634) x US-08-723-624-18 (1-3035)
 QY 16 SerLeuGlyThrProAlaAlaArgHisPheProArgAsnGluMetThrGlnHisGluGln 35
 DB 1176 GCGGTGCGACGACGCGCTGAGAGATCCCGGCTGTCACTCC----- 1220
 QY 36 ProLeuIleValArgProGlnArgThrSerSerArgGluLeuValAsnLeuAspGly 55
 DB 1221 -----CTTCTGTGATGCTCTGTGAACCCCAACCGTGAATACTCAAAAACTCGACGCG 1274
 QY 56 LeuTrpLeuPheAlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnProTrp 72
 DB 1275 CTGTGGGCAATTGATGCTGATGCGCAAACTGTGAATGATGAG-----CGTGGGTGG 1328

QY		73	ThrlAProLeuPseProLyGlyLeuGlucysProValProAlaSerTyrAsnAspIle	92
Db		1329	GAAAGCGCTTTACAGAAGAACCCGGCAATTGCTGTGCCAGGCAGTTTTAACATCAGTTC	1388
QY		93	IleSerArgJulienHisAspHisValGlyTrrValTyrTyrGlnArgGluValIleVal	112
Db		1389	GCCCATGCAGATATTGCTGAATTATTCGGGGCAACGCTGTGATTCAGCGGAAGTCTTTATA	1448
QY		113	ProLySGlyTrpSerGlnGluArgGlyLeuValArgAlaGluSerAlaThrHisHicGly	132
Db		1449	CGGAAGGTTGGGCGCCAGGCGATTCGTGGCGCTTTCGATCGCGTCACTCATATTACGGC	1508
QY		133	ArgIleTyrValAsnAsnArgLeuValAlaGlnHisValGlyGlyTyrThrProPhcGlu	152
Db		1509	AAAGTGCTGGTCATTAATCACGAAGTAGTAGCATCAGGGCGGGCTATACCATTTCGA	1568
QY		153	ALAAspValThrGluLeuValAlaProGlyGluysPheIleArgLeuThrIleGlyValAsn	172
Db		1569	GCCCATGTCACGCGGTATGTTATTCGCCGGAAAACTGACTATTCACCGTTGTGTGAAC	1628
QY		173	AsnGluLeuThrHisGlyThrIleProProGlyLysIleThrThrGlyAsnAlaThrGly	192
Db		1629	AACCAACTGAACCTGGCAGACTATCCCCCGGGAATGGTGATTACC--GACGAAAACGCC	1685
QY		193	LysArgIleGlnThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaArgSerIle	212
Db		1686	AAGAAAAAGCAGCTCTTACTCTCATGATTTCTTTAACTTATGCCGGGATCATCGCAGCSTA	1745
QY		213	TrrLeuTyrSerValProGlnGlnHisIleGlnAspIleThrValIleThrAspVal---	231
Db		1746	ATGCTCTACACACGCCGCCACAACCTGGGGAGACGATACCGTGTACCGAGTCCGC	1805
QY		232	--AspGlyAspAsnGlyLeuIleAsnTyrGluValGluValAlaAsnGlnThrThrGly	250
Db		1806	CAAAGCTGTAAACACGCGCTGTGTTACTGGCGAGGG--GTGGCCAAT-----GGT	1853
QY		251	GlnIleGlnIleSerValIleAspGlnAspGlyAlaIleValAlaLysAlaSerGlyValA	270
Db		1854	GATGTCAACCGTTGAACCTCGATATCCGATCAACAGGTGTTCGAACGTGACAAAGCACT	1913
QY		271	GlnGlyTrrValIleThrIleProSerValLysLeuTrrPgnProGlyValAlaIleTyrLeuTyr	290
Db		1914	AGCGGCACTTTCGAAGTGTGAATCCGACCTTCGCAACCGGGGAAGCTTATCTCAT	1973
QY		291	GlnLeuGlnValAsnIleValIleGlySerSerGlyAspValIleAspThrTyrAsnLeuAla	310
Db		1974	GAACGTGTCGTC---ACAGCCAAAAGCCAGACAGATGTGATATCTACCCGCTTGGC	2027
QY		311	ThrGlyValArgThrValIleValAlaGlySerGlnPheLeuIleAsnGlyLysProPhe	330
Db		2028	GTCCGCACTCCGCTGTGCGTSCAGTGAAGGGCAACGTTCTCGATTAAACCAAAACCGTTC	2087
QY		331	TyrPheThrGlyPheGlyLysHisGlnAspThrAlaValArgGlyLysGlyHisAspPro	350
Db		2088	TACTTTACTGCTTGTGTCGTCAATGATGATGGGCACTTGGCGGCAAGATTCGATAC	2147
QY		351	AlaIleTrrMetValHisAspPheGlnLeuMetLysTrrIleGlyAlaAsnSerPheArgThr	370
Db		2148	GTGCTGAGGTCACAGACACGACATTAATGACGTGATTTGGGGCCCACTCTCCAGTACC	2207
QY		371	SerHisTyrProTyrAlaGlnGluValMetAspPheAlaAspArgAsnGlyIleValVal	390
Db		2208	TTCGATTACCTTACGTCGAAGAKTGTCTCACTCGGCAAGATGAACATGCGATCTGTGTG	2267
QY		391	IleAspGlnThrProAlaValGlyLeuAsnIleAlaLeu--MetGlyValSerGlnSer	409
Db		2268	ATTGATGAATACGTGCTGTGGCTTTAACTCTTTAAGCATGTGTTGGAAGCGGCG	2327
QY		410	GlyAlaPro---GlnThrPheThrProAspAlaIleAsnAspLysThrGlnGluAlaHis	428
Db		2328	AACAAAGCCGAAGAACTGTACGCGAAGAGCTCAAAGGGGAAAACTCAGCAAGCGCAC	2387

```

OY 429 LygGlnIaIaIaAGGluIeIIaIaAArgAspLyAsnHlgaIaSerValaIaMetTrp 448
Db 2388 TTACAGGGCGATTAAAGAGCTGATATGCGCGTACAAAAACCAACCCAGCGTGTATGTGCG 2447
OY 449 SerIleAlaAngIuPProIaIaSerHleGluAspGluIaIaArgGluTrpPheGluProIeu 468
Db 2448 AGTATTTGCGAACGAAACCGGATACCGGTCCGCAAGGTGCACGGGAATATTTTCGCCCACTG 2507
OY 469 ThrAsnLeuThraGgInLeuAspProThraGProIleThraPheAlaSerValaIcIYthr 488
Db 2508 GCGGAGACGACCGCTTAACCTGACCCGACCGGTCCGATCACCTCCGTCGAAGTATGTTC 2567
OY 489 AlaThrTrpGlnLeuAspArgIleSerAspLeuPheAspValaSerCysIleAsnArgTYr 508
Db 2568 TGGCAGCGCTCACACCGATCCATCAGGAGATCTTTGATGTGTGTGCTGTACACCGTTAT 2627
OY 509 PheGluTrpTrpSerGlnThrGlyAspLeuGluIuIaGluAlaAlaIeugIuIaGlu 528
Db 2628 TACGAGATGATATGTCCAAAGCGCGGATTTGGAAACGGCAGAAAGTAACTGGAAAAAGAA 2687
OY 529 LeuHlscglTYrTrpGlnGluIuIaPheHleArgProIleValaIaMetThrGluTYrGlyAlaAsp 548
Db 2688 CTTCTGGCGCTGGCAGGAGAAATCTCATCAGCCGATTTATCAACACCGAATACGGCGTGGAT 2747
OY 549 ThrIleuAlaGlyIeAsnHlserIleLeuGluIleuProIrrpSerGlnIuIaPheGlnValaIa 568
Db 2748 ACGTTAGCCGGCGCTGCACCTCATATACACCGACATGTGGAGTGAAGATATCACTGTGCA 2807
OY 569 MetIleuAspMetTYrHlIaArgValaIaPheAspArgIleGluSerMetAlaGluIuHleVala 588
Db 2808 TGGCGTGGATATGATATCACCGCGTCTTTGATGTGCTCAGCGCCGTCGTCGCGTGAACAGTAA 2867
OY 589 TrpAsnPheAlaAspPheGlnThrAsnLeuGluIleIleArgValaAspGlyAsnIuIaGly 608
Db 2868 TGGAAATTCGCCGCAATTTTTCGCACCTCCGACGAGCATATTTGCCGCTTGGCGGTAAACAGAA 2927
OY 609 GlyAlaPheThrArgAspArgLyProIaIaIaIaHlserIleuArgAlaArgTYr 628
Db 2928 GGGATCTTCATCCCGACCGCAACCGAAGTCGGCGCTTTCTGTGCAAAAAACGCTGG 2987
OY 629 ThrSerIleAsp 632
Db 2988 ACTGGCATGAAC 2999

RESULT 5
US-08-723-624-19
; Sequence 19, Application US/08723624
; Patent No. 5861377
; GENERAL INFORMATION:
; APPLICANT: Rose, Alan B.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ENHANCING
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ENHANCING
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: ARNOLD, WHITE & DUNKEE
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/723,624
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165

```

REFERENCE/DOCKET NUMBER: BTIP:002
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 512/418-3000
 TELEFAX: 512/474-7577
 INFORMATION FOR SEQ ID NO: 19:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3824 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-723-624-19

Alignment Scores:
 Pred. No.: 5,56e-174 Length: 3824
 Score: 1711.00 Matches: 330
 Percent Similarity: 70.50% Conservative: 93
 Best Local Similarity: 55.00% Mismatches: 161
 Query Match: 51.01% Indels: 16
 Gaps: 9

US-10-757-093-4 (1-634) x US-08-723-624-19 (1-3824)

```

QY 40 ValArgProGlnArgThrSerSerArgGluLeuValAsnLeuAspGlyLeuTrpLysPhe
DB 2016 TTAAGTCTGCTAGAAACCCCAACCCGGAATCAAAAACCTGCGGCGCTGCGGCGATTC
QY 60 AlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnProTrpPthAlaProLeu
DB 2076 AGCTGTGATCGCAAAAACCTGTGAAATTGATCAG-----CGTTGGTGGAAACGCGCTTA
QY 77 ProLysGlyLeuGlyCysProValProAlaSerTyraAsnAspIlePheIleSerArgLys
DB 2130 CAAGAAACCGCGCAATTCTGTGCGCAGCACTTTTAAACATCACTTCCCGCATCGCAT
QY 97 IleHisAspHisValGlyTrpValTyrTyrGlnArgGluValIleValProLysGlyTrp
DB 2190 ATTGTAATATATCGGCGCAACGCTGTGATCAGCGGCAAGCTTTATACCGAAAGGTTCG
QY 117 SerGlnGluArgTyrLeuValArgAlaGluSerAlaThrHisHisGlyArgIleTyrVal
DB 2250 GCGAGCGAGGTATCGTGTGCTGCTTCGATCGAGTCTCATTAACGCAAAAGTGTGTC
QY 137 AsnAsnArgLeuValAlaGluHisValGlyGlyTyrThrProPheGluAlaAspValThr
DB 2310 AATTAATCAAGATGATGAGCATCAAGGCGCTAATGCGCATTTGAAGCCGATGTCACG
QY 157 GluLeuValAlaProGlyGluLysPheArgLeuThrIleGlyValAsnAsnGluLeuThr
DB 2370 CCGTATGTTATGCGCGGAAAGTGAATGATACCGTTGTTGTGAAACAAGCACTGAAC
QY 177 HisGluThrIleProProGlyLysIleThrThrGlyAsnAlaThrGlyLysValGlnIleGln
DB 2430 TGGCAGACTATCCCGCGGGAATGCTGATTACC---GACGAAAAACGCAAAAGAAAGCAG
QY 197 ThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaArgSerIleTrpLeuTyrSer
DB 2487 TCTTACTTCATATTTCTTAATCTATGCGGAATTCATGCAAGCGTAATGCTCTACACC
QY 217 ValProGlnGlnHisIleGlnAspIleThrValValThrAspVal-----AspGlyAsp
DB 2547 ACCCGCAACACCGGCGGAGCATATCACCGTGTGACGATCGCGCAAGCACTGAAC
QY 235 ArgGlyLeuIleAsnTyrGluValAlaGluValAlaAsnGlnThrThrGlyGlnIleGlnIle
DB 2607 CAGCGCTGCTGTACTGCGCAGGTG---GTGCGCAAT-----GGTGAATGTCAGCGTT
QY 255 SerValIleAspGluAspGlyAlaIleValAlaLysAlaSerGlyAlaGlnGlyThrVal
DB 2655 GAACTGGTGTAGCGGATCAACAGGTGTGCACTGGAACAGGCACTACCGGACCTTTC
QY 275 ThrIleProSerValLysLeuTrpGlnProGlyValAlaIleTyrLeuTyrGlnLeuGlnVal
DB 2715 CAAGTGTGTAATCCGACCTCTGCGCAACCGGCTGAAGATTATCTCTATGAACGTGTGCTC
  
```

```

QY 295 AsnIleValGlySerSerGlyAspValValAspThrTyrAsnLeuAlaThrGlyValArg
DB 2775 -----ACAGCCAAAAGCCAGACAGATGTGATTAATCAACCCGCTTCCGCGCATCCGG
QY 315 ThrValLysValAlaGlySerGlnPheLeuIleAsnGlyLysProPheTyrPheThrGly
DB 2829 TCAGTGGCAGTGAAGGCGCAACAGTTCCTGATTAACACAAACCGTTCTACTTATTCGCG
QY 335 PheGlyLysHisGluAspThrAlaValArgGlyLysGlyHisAspProAlaTyrMetVal
DB 2889 TTTCGTGCTATGAATGCGAGCTGCGCGCAAGATTCAATACGTGCAATGTCG
QY 355 HisAspPheGlnLeuMetLysTrpIleGlyAlaAsnSerPheArgThrSerHisTyrPro
DB 2949 CACGACCAACGCAATTAATGACATGATTTGGGCCAATCTCTACCTGATCTGCACTTACCT
QY 375 TyrAlaGluGluValMetAspPheAlaAspArgAsnGlyIleValValIleAspGluThr
DB 3009 TACGCTGAAGATGCTCGCATGCGCAGATGAACATGCACTGCTGTGATTTGAACACT
QY 395 ProAlaValGlyLeuAsnIleAlaLeu---MetGlyValSerGluSerGlyAlaPro---
DB 3069 GCTGCTGTGCGCTTTAACTCTTTAGGCAATGTTTCGAAACCGGCAACAACCGGAAA
QY 413 GlnThrPheThrProAspAlaIleAsnAspLysThrGlnGluValHisLysGlnAlaIle
DB 3129 GAACTGTACAGCAAGAGGAGTCAACGGGCAACCTCAGACCGCACTTACAGCGCATT
QY 433 ArgGluLeuIleAlaArgAspLysAsnHisAlaSerValValMetTrpSerIleAlaAsn
DB 3189 AAGAGCTGATATACGGGTGACAAACCAACCAAGGCTGTGATGTGGAATATTCGCAAC
QY 453 GluProAlaSerHisGluAspGlyAlaArgGluTyrPheGluProLeuThrAsnLeuThr
DB 3249 GAAACCGATACCCGTCGCAAGGTGCAACGGAATATTTCCGCGCACCTGCGGGAACACG
QY 473 ArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGlyThrAlaThrTyrGln
DB 3309 CGTAACTGACACCGACGCGCTCGATCACTGCTCATATTAATGTTCTCCGACCTCAC
QY 493 LeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArgTyrPheGlyTrpTyr
DB 3369 ACCGATACCATCAGCATCTTGTGATGTCGTGCTGCAACCGTTATTCGATGCTAT
QY 513 SerGlnThrGlyAspLeuGluAlaGluAlaIleGluLysGluLeuHisGlyTrp
DB 3429 GTCCAAAAGCGCGATTTGGAAACGCGACAGAAAGTACTGCAAAAGAACTTCTGCGCTCG
QY 533 GlnGluLysPheHisArgProIleValMetThrGluTyrGlyValAlaAspThrLeuAlaGly
DB 3489 CAGAGAAATGATGATCAACCGATTAATCATCAGGAATACGCGCTGTGATACGTTAGCCGGG
QY 553 LeuHisSerIleLeuGlyLeuProTrpSerGluGluPheGlnValGlnMetLeuAspMet
DB 3549 CTGCACTCAATGATACACCGACATGTGAGTGAAGATATCAAGTGTGATGCTGATATG
QY 573 TyrHisArgValPheAspArgIleGluSerMetAlaGlyGluHisValIleTrpAsnPheAla
DB 3609 TATCACCGCGCTTTGATCGCTGACGCGCTGTGCTGTGTAACAGGATATGAATTTCCGCC
QY 593 AspPheGlnThrAsnLeuGlyIleIleArgValAlaAspGlyValAsnLysGlyValPheThr
DB 3669 GATTTTGCAGCTCTCGAAGGCAATATTTGCGGTGTGCGGTAAACAAGAGGATCTTCACT
QY 613 ArgAspArgLysProLysAlaAlaHisSerLeuArgAlaArgTrpThrSerIleAsp
DB 3729 CCGGACCGCAACCGAAGTGTGCGCGCTTTCTCTCTCAAAAACGCTGAGATGCGCATGAAC
  
```

RESULT 6
 US-09-893-525-36
 ; Sequence 36, Application US/09893525
 ; Patent No. 6753167

```

GENERAL INFORMATION:
APPLICANT: Moloney, Maurice M.
TITLE OF INVENTION: Preparation of Heterologous Proteins on Oil Bodies
FILE REFERENCE: 9369-172
CURRENT APPLICATION NUMBER: US/09/893,525
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 09/210,843
PRIOR FILING DATE: 1998-12-15
PRIOR APPLICATION NUMBER: US 08/846,021
PRIOR FILING DATE: 1997-04-25
PRIOR APPLICATION NUMBER: US 08/366,783
PRIOR FILING DATE: 1994-12-30
PRIOR APPLICATION NUMBER: US 08/142,418
PRIOR FILING DATE: 1993-11-16
PRIOR APPLICATION NUMBER: US 07/659,835
PRIOR FILING DATE: 1991-02-22
NUMBER OF SEQ ID NOS: 42
SOFTWARE: PatentIn version 3.1
SEQ ID NO 36
LENGTH: 4652
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Phae-GUS-phas
NAME/KEY: CDS
LOCATION: (1548)..(3359)
OTHER INFORMATION:
US-09-893-525-36

Alignment Scores:
Pred. No.: 1,25e-173 Length: 4652
Score: 1709.00 Matches: 338
Percent Similarity: 67.34% Conservative: 97
Best Local Similarity: 52.32% Mismatches: 179
Query Match: 50.95% Indels: 32
Gaps: 11

US-10-757-093-4 (1-634) x US-09-893-525-36 (1-4652)

QY 9 LeuLeuSerLeuAlaAlaProSerLeuGlyThrProAlaAlaArgHisPheProAlaGln 28
DB 1419 CTTCTCTCTATATATATACCT---ATAAATACCTCTAATATCACTCTTTCATCAT 1475
QY 29 GluMetThrGlnHisGluGlnProLeuIle----- 38
DB 1476 CCATCAATCCAGAGTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 1535
QY 39 -----LysValArgProGlnArgThrSerSerArgGluLeuValAsnLeu 53
DB 1536 TACTACTACCAATGAGTCTTACGCTCTGTAGAAACCCCAACCGGTGAATCAAAAAAAG 1595
QY 54 AspGlyLeuThrLysPheAlaLeu-----AlaSerGlyLeuAsnSerThrAlaGln 70
DB 1596 GACGGCTGTGGGCAATTCAGTCTGATCGCGAAACCTGTGGAATGATGTCG-----CGT 1649
QY 71 ProThrThrAlaProLeuProLysGlyLeuGluCysProValProAlaSerTyraAsnArg 90
DB 1650 TGGTGGGAAAGCCGCTTACAAAGAACCGGCAATTCGTGTGTGACAGGAGTTTAAACGAT 1709
QY 91 IlePheIleSerArgGluIleHisAspHisValGlyTyrValTyrTyrGlnArgGluVal 110
DB 1710 CAGTTGGCGGATGCAATATATGTAATATAGCGGGCAAGCTCTGATACAGCGCAAGTC 1769
QY 111 IleValProLysGlyTyrPheSerGlnGluArgTyrLeuValArgAlaGluSerAlaThrHis 130
DB 1770 TTATATCCCAAGAGTTGGGCAAGCCGACGATGCTGTGCTGCTTTCGATGCGGTCACTCAT 1829
QY 131 HisGlyArgIleTyrValAsnAsnArgLeuValAlaGlnHisValGlyTyrThrPro 150
DB 1830 TACGGCAAGAGTGTGGGTCAATATCAAGGAAGATGATGACATGACGCGGCTATACGCCA 1889
QY 151 PheGluAlaAspValThrGluLeuValAlaProGlyGluLysPheArgLeuThrIleGly 170

```

```

DB 1890 TTGAAAGCCGATGTACCGCGATGTTATTCGCGGAAAGGTAGTATACCGTTGCT 1949
QY 171 ValAsnArgGluLeuThrHisGluThrIleProProGlyLysIleThrThrGluAsnAla 190
DB 1950 GTGAACAAAGAACTGAATCTGAGACTATCCCGCCGGAATGTGATTACC---GACGAA 2006
QY 191 ThrGlyLysArgIleGlnThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaArg 210
DB 2007 AACGGCAAGAAAAGAGCTTACTTCCATGATTTCTTAACTATGCCGAATCCATCCG 2066
QY 211 SerIleThrLeuTyrSerValProGlnGlnHisIleGlnAspIleThrValAlaThrAsp 230
DB 2067 AGCGTAACTCTACACCAACCGCAACACCTGGGTGACGATATCACCGTGGACGAT 2126
QY 231 Val-----AspGlyAspAsnGlyLeuIleAsnTyrGluValGluValAlaAsnGlnThr 248
DB 2127 GTCCGCGAAGACTGTAAACACCGCTGTGTGCTGGCAGGTG---GTGGCCAT----- 2177
QY 249 ThrGlyGlnIleGlnIleSerValIleAspGlyAspGlyAlaIleValAlaLysAlaSer 268
DB 2178 ---GGTGAATGACGCTTAACCTGCTGATGCGGATCAACAGGTGTTGCACTGACAA 2234
QY 269 GlyAlaGlnGlyThrValThrIleProSerValLysLeuThrGlnProGlyAlaAlaTyr 288
DB 2235 GGCATAGCGGGACCTTGGCAAGTGTGAATCCGACCTCTGGCAACCGGTGAAGGTAT 2294
QY 289 LeuTyrGlnLeuGlnValAsnIleValGlySerSerGlyAspValValAspThrTyrAsn 308
DB 2295 CTTATGAACGTGGCTC-----ACAGCCAAAGACAGACAGAGTGTATTTCAACCG 2348
QY 309 LeuAlaThrGlyValArgThrValLysValAlaGlySerGlnPheLeuIleAsnGlyLys 328
DB 2349 CTTGCGGTGGGATCCGGTCACTGAGTGGAGGAGGCAACAGTCTGTATTAACCAAA 2408
QY 329 ProPheTyrPheThrGlyPheGlyLysHisGluAspThrAlaValArgGlyLysGlyHis 348
DB 2409 CGGTCTACTTAACTGCTGTGGTGTGCTGATGAAGATGCGGACTTACGTGGCAAGAGATT 2468
QY 349 AspProAlaTyrMetValHisAspPheGlnLeuMetLysTyrIleGlyLysAlaAsnSerPhe 368
DB 2469 GATTAACGTCTGATGTGACGACACCGCATTAATGATGATGGAGTGGGCAACTCTTAC 2528
QY 369 ArgThrSerHisTyrProTyrAlaGluGluValAlaMetAspPheAlaAspArgAsnGlyIle 388
DB 2529 CGTACCTCCATTAACCTTACCTGATGAAGATGCTGATGCGGCATGATGAACATGGCATC 2588
QY 389 ValValIleAspGluThrProAlaValGlyLeuAsnIleAlaLeu---MetGlyValSer 407
DB 2589 GTGGTATTGATGAACCTGCTGCTGCGGCTTTTCCCTCTTAAAGCATTTGGTTCCGA 2648
QY 408 GluSerGlyAlaPro---GlnThrPheThrProAspAlaIleAsnAspLysThrGlnGlu 426
DB 2649 GCGGGCAACAGCCGAAAGAACTGTACAGGAGGAGCTCAACGGGAAACTCCGCAA 2708
QY 427 AlaHisLeuGlyAlaIleArgGluLeuIleAlaArgAspLysAsnHisAlaSerValVal 446
DB 2709 GCGCACTTACAGCGCATTTAAAGCTGATAGCGCGGACAAACCAACCAACCGGTGCTG 2768
QY 447 MetTrpSerIleAlaAsnGluProAlaSerHisGluAspGlyAlaArgGluTyrPheGlu 466
DB 2769 ATGTGAGATATGCCAACAGACCGGATACCCGCCGCAAGTGACCGGAAATATTTCGGG 2828
QY 467 ProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnVal 486
DB 2829 CCACTGCGGAAAGCAACGGGTAACTCGACCGGACCGGTCAATCACTCGGTCAATGTGA 2888
QY 487 GlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPheAspValSerCysIleAsn 506
DB 2889 ATGTTTGGAGGCTTACACCGATATCAAGGATCTTTGATGATGCTGTGCTGCAAC 2948
QY 507 ArgTyrPheGlyTyrPyrSerGlnThrGlyAspLeuGluAlaGluAlaAlaLeuGlu 526

```


Db 2949 CGTTATTACGGATGTATGTCCAAACGGCGGATTTGGAAACGGACAGAAAGTACTGAA 3008
 QY 527 LysGluLeuHisGlyTrpGlnGluValPheHisArgProIleValMetThrGluTyrGly 546
 Db 3009 AAGAACTCTGCGCTGGCAGGAGAACTGCAATGACGATATCATCACCGAAATACGGC 3068
 QY 547 AlaSerThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTrpSerGluGluPheGln 566
 Db 3069 GCGGATACGTTACCGGGGCGCACTCATATACACGACATGAGATGAAGATACG 3128
 QY 567 ValGluMetLeuAspMetTyrHisArgValPheAspArgIleGluSerMetAlaGlyGlu 586
 Db 3129 TGTGATGCGTGGATATGATACACCGCTCTTGATGCGCTGACGCCGCTGCGGAGAA 3188
 QY 587 HisValThrAspPheHisAspPheGlnThrAsnLeuGlyIleIleArgValAspGlyAsn 606
 Db 3189 CAGGATAGGAATTCCTCCGATTTTGGACCTCGCAAGGCATATGTGCGCTGGCGGTAAC 3248
 QY 607 LysLysGlyValPheThrArgAspArgLysPheGlyAlaIleAlaHisSerLeuArgAla 626
 Db 3249 AAGAAAGGAGTCTTCACTCGCGACCCGAAACCGAAGTCGGCGCTTTTCTGCTGCAAAA 3308
 QY 627 ArgTrpThrSerIleAsp 632
 Db 3309 CGCTGACTGCGCATGAAAC 3326
 RESULT 7
 US-09-893-525-41
 / Sequence 41, Application US/09893525
 / Patent No. 6753167
 / GENERAL INFORMATION:
 / APPLICANT: Moloney, Maurice M.
 / APPLICANT: Van Rooijen, Gijze
 / TITLE OF INVENTION: Preparation of Heterologous Proteins on Oil Bodies
 / FILE REFERENCE: 9369-117
 / CURRENT APPLICATION NUMBER: US/09/893,525
 / PRIOR FILING DATE: 2001-06-29
 / PRIOR APPLICATION NUMBER: US 09/210,843
 / PRIOR FILING DATE: 1998-12-15
 / PRIOR APPLICATION NUMBER: US 08/846,021
 / PRIOR FILING DATE: 1997-04-25
 / PRIOR APPLICATION NUMBER: US 08/366,783
 / PRIOR FILING DATE: 1994-12-30
 / PRIOR APPLICATION NUMBER: US 08/142,418
 / PRIOR FILING DATE: 1993-11-16
 / PRIOR APPLICATION NUMBER: US 07/659,835
 / PRIOR FILING DATE: 1991-02-22
 / NUMBER OF SEQ ID NOS: 42
 / SOFTWARE: Patent version 3.1
 / SEQ ID NO 41
 / LENGTH: 5390
 / TYPE: DNA
 / ORGANISM: Artificial Sequence
 / FEATURE:
 / OTHER INFORMATION: phase-caleo-GUS-phas
 / NAME/KEY: CDS
 / LOCATION: (1548)..(4097)
 / OTHER INFORMATION:
 / US-09-893-525-41
 Alignment Scores:
 Pred. No.: 3 34e-173 Length: 5390
 Score: 1706.00 Matches: 329
 Percent Similarity: 70.50% Conservative: 94
 Best Local Similarity: 54.83% Mismatches: 161
 Query Match: 50.86% Indels: 16
 Gaps: 9
 US-10-757-093-4 (1-634) x US-09-893-525-41 (1-5390)
 QY 40 ValArgProGlnArgThrSerSerArgGluLeuValAsnLeuAspGlyLeuTrpLysPhe 59
 Db 2292 TTAACGTCCTGTAGAAACCCCAACCCGCTGAATAAATACTCAACGCGCTGTGGCATTC 2251

QY 60 AlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnProTrpThrAlaProLeu 76
 Db 2352 AGCTGATCGCGAAACTGTGAAATGATCAG-----CGTTGCTGGAGAAACGGCTTA 2405
 QY 77 ProLysGlyLeuGluCysArgProValProAlaSerTyrAsnAspIlePheIleSerArgGlu 96
 Db 2406 CAAGAAAGCCGGCAATTCCTGTGCCAGGACAGTTTAAAGATCAGATCCGCCAGTACGAT 2465
 QY 97 IleHisAspHisValGlyTrpValTyrTyrGlnArgGluValIleValProLysGlyTrp 116
 Db 2466 ATTCTGAATTAATTCGGGCAACGCTGTATACCGGGAAGTCTTATTACGGAAGGTGG 2525
 QY 117 SerGlnGluArgTyrLeuValArgAlaGluSerAlaThrHisHisArgIleTyrVal 136
 Db 2526 GCAGGCCAGGATATGCTGCTGCTTCATGCGGTACCTCATTAACGCAAGTGGGTC 2585
 QY 137 AsnAsnArgLeuValAlaGluHisValGlyTyrThrProPheGluAlaAspValThr 156
 Db 2586 AATPATCAGGAAGATGATGAGCATCAGCGCGGTATACGCATTTGAAGCCGATGTCAGC 2645
 QY 157 GluLeuValAlaProGlyGluLysPheArgLeuThrIleGlyValAsnAsnGluLeuThr 176
 Db 2646 CGGTATGTTATTCGCGGAAAGTGTACATCACCGTTGTGTGAACAACGAATGAAAC 2705
 QY 177 HisGluThrIleProProGlyLysIleThrThrGlyAsnAlaThrGlyLysArgIleGln 196
 Db 2706 TGGCAGACTATCCCGCGGAGATGTGATTAAC--GACGAAACGCAAGAAACGACG 2762
 QY 197 ThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaArgSerIleTrpLeuTyrSer 216
 Db 2763 TCTTACTTCATGATTTCTTTAACTATGCGGAATCATCGACGCGTAATGCTTACACC 2822
 QY 217 ValProGlnGlnHisIleGlnAspIleThrValIleThrAspVal-----AspGlyAsp 234
 Db 2823 ACCCGCAACACCTGCGTGCAGCATATCACCGGTGTACCATCTCCGCAAGCTGTAC 2882
 QY 235 AsnGlyLeuIleAsnTyrGluValGluValAlaAsnGlnThrGlyGlnIleGlnIle 254
 Db 2883 CACGCGCTGTGTACCTGGCAGGTG--GTGGCAAT-----GATATGTCAAGCTT 2930
 QY 255 SerValIleAspGluAspGlyAlaIleValAlaLysAlaSerGlyAlaGlnGlyThrVal 274
 Db 2931 GAACGCGATGATCGGATCAACAGCGTGTGCAACGTGCAAGGACGCTAGCGGAGCTTGG 2990
 QY 275 ThrIleProSerValLysLeuTrpGlnProGlyAlaIleTyrLeuTyrGlnLeuGlnVal 294
 Db 2991 CAAGTGTGAATCCGCACTGTGCAACCGGGTGAAGTATCTTATGAAACTGTGCTGC 3050
 QY 295 AsnIleValGlySerSerGlyAspValValAspThrTyrAsnLeuAlaThrGlyValArg 314
 Db 3051 -----ACAGCCAAAGCCAGACAGAGTGTGATATCTACCCGCTGCGCATCCG 3104
 QY 315 ThrValLysValAlaGlySerGlnPheLeuIleAsnGlyLysProPheTyrPheThrGly 334
 Db 3105 TCAGTGCAGTGAAGGCCAACAAGTTCCTGATTAACCAAAACGCTTACTTATCTGGC 3164
 QY 335 PheGlyLysHisGluAspThrAlaValArgGlyLysGlyHisAspProAlaTyrMetVal 354
 Db 3165 TTGGTGTGTCATGAGATGCGGACTTACGTCGCAAGGATTCGATTAACGTGCTGATGTG 3224
 QY 355 HisAspPheGlnLeuMetLysTrpIleGlyAlaAsnSerPheArgThrSerHisTyrPro 374
 Db 3225 CACGACACGACATTAATGACATGAGTTGGGCCCAACTCTTACCGTATCCCTGCAATCCCT 3284
 QY 375 TyrAlaGluGluValMetAspPheAlaAspArgAsnGlyIleValValIleAspGluThr 394
 Db 3285 TACGCTGAAGAGATGTCATCGCGCAATGCAAGATGCAATGCTGTGATTAAGTAACT 3344
 QY 395 ProAlaValGlyLeuAsnIleAlaLeu--MetGlyValSerGluSerGlyAlaPro--- 412
 Db 3345 GCGTGTGCGGCTTTTCCGCTCTTATGAGCATTTGTCGAAGCGGCAACGAGCCGAA 3404

```

QY 413 GlnThrPheThrProAspAlaIleAsnAspLysThrGlnGluValHisIleSerGlnAlaIle 432
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 3405 GAACCTACAGCAGACAGGACGACGAGGAACTCAGCAAGCGCATTTACAGGGCATT 3464
QY 433 ArgGluLeuIleAlaArgAspLysAsnHisAlaSerValIleMetTrpSerIleAlaAsn 452
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 3465 AAAGAGCTGATAGCGCGTGCACAAAACCAACCCAAAGCGTGATGTGAGTATTCGCAAC 3524
QY 453 GluProAlaSerHisSerGluAspGluValIleArgGluTyrPheGluProLeuThrAsnLeuThr 472
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 3525 GAACCGGATACCGCTCGCAGAGGTGCACGGAAATATTTCCGCCACTGCGGAGCAAG 3584
QY 473 ArgGluLeuAspProThrArgProIleThrPheAlaAsnValGlyThrAlaThrTyrGln 492
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 3585 CGTAACTGACCCGACCGCGCTCGATCCTCGTGCTAAATGTATGTTCTTCGCAAGCTCAC 3644
QY 493 LeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArgTyrPheGlyTyr 512
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 3645 ACCGATACCATCAGCATCTCTTGATGTGCTGTGCTGACCAACGTTATTAAGGATGTAT 3704
QY 513 SerGlnThrGlyAspLeuGluGluValIleGluValAlaLeuGluLeuLeuHisGlyTyr 532
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 3705 GTCCAAAGCGCGCATTTGAAACGCAAGAGTACTGGAAAAAGAACTTCGGCCCTGG 3764
QY 533 GlnGluLeuPheHisArgProIleValIleMetThrGluTyrGlyValAspThrLeuAlaGly 552
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 3765 CAGGAGAACTGCTATCAGCCGATTCATCCAGATACCGCGCTGGATACGTCGCGG 3824
QY 553 LeuHisSerIleLeuGlyLeuProTyrSerGluGluPheGlnValGlnMetLeuAspMet 572
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 3825 CTGCACATCATATGACACCCAGCATGTGAGTGAAGATACAGTGCATGCTGCGATATG 3884
QY 573 TyrHisArgValPheAspArgIleGluSerMetAlaGlyGluHisValIleThrAsnPheAla 592
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 3885 TATCACCGGCTCTTATGTCGCGACCGCGCTGCTGCGGACAGATATGAAATTTTCGCC 3944
QY 593 AspPheGlnThrAsnLeuGlyIleIleArgValAspGlyAsnLysGlyValPheThr 612
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 3945 GATTTTGCACCTCCGACAGGCATATTCGCGTTGGCGGTACCAAGAAAGGATCTTCACT 4004
QY 613 ArgAspArgLysProLysAlaAlaAlaHisSerLeuArgAlaArgTyrThrSerIleAsp 632
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 4005 CGCGACCGCAACCGAAGTCGCGCTTTCTGCTCAAAAACGTCGACATGCAATGAC 4064

RESULT 8
US-09-893-525-38
; Sequence 38, Application US/09893525
; Patent No. 6753167
; GENERAL INFORMATION:
; APPLICANT: Moloney, Maurice M.
; TITLE OF INVENTION: Preparation of Heterologous Proteins on Oil Bodies
; FILE REFERENCE: 9369-172
; CURRENT APPLICATION NUMBER: US/09/893,525
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 09/210,843
; PRIOR FILING DATE: 1998-12-15
; PRIOR APPLICATION NUMBER: US 08/846,021
; PRIOR FILING DATE: 1997-04-25
; PRIOR APPLICATION NUMBER: US 08/366,783
; PRIOR FILING DATE: 1994-12-30
; PRIOR APPLICATION NUMBER: US 08/142,418
; PRIOR FILING DATE: 1993-11-16
; PRIOR APPLICATION NUMBER: US 07/659,835
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 38
; LENGTH: 5418
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: phas-oleo GUS--phas

```

```

; NAME/KEY: CDS
; LOCATION: (1555)..(1908)
; OTHER INFORMATION:
; NAME/KEY: CDS
; LOCATION: (2149)..(4125)
; OTHER INFORMATION:
US-09-893-525-38

Alignment Scores:
Pred. No.: 3,36e-173 Length: 5418
Score: 1706.00 Matches: 329
Percent Similarity: 70.50% Conservative: 94
Best Local Similarity: 54.83% Mismatches: 161
Query Match: 50.86% Indels: 16
DB: 4 Gaps: 9

US-10-757-093-4 (1-634) x US-09-893-525-38 (1-5418)
QY 40 ValArgProGlnArgThrSerSerArgGluLeuValAsnLeuAspGlyLeuTyrPhe 59
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2320 TTACGTCCTGTGAAACCCCAACCGGTGAATCAAAAACCTGACGCGCTGTGGCATTC 2379
QY 60 AlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnProTyrThrAlaProLeu 76
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2380 AGCTCGATCGCGAAACCTGTGAATGTATCAG-----CGTGTGGGAAAGCGGTTA 2433
QY 77 ProLysGlyLeuGluCysProValProAlaSerTyrAsnAspIlePheIleSerArgLys 96
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2434 CAGAAAGCCGGGCAATTCTGTGCGAGCATTTTAAAGCATGTTGCCCATGCAGAT 2493
QY 97 IleHisAspHisValGlyTyrValTyrTyrGlnArgGluValIleValProLysGlyTyr 116
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2494 ATTGCAATATTAACGGGCAACGTCGTGATCAGCGGAAAGTCTTATACGAAAGTTGG 2553
QY 117 SerGlnGluArgTyrLeuValArgAlaGluSerAlaThrHisHisGlyArgIleTyrVal 136
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2554 GCAAGCCACCGTATGTCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2613
QY 137 AsnAspArgLeuValAlaGluHisValGlyGlyTyrThrProPheGluValAspValThr 156
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2614 AATATACAGAAAGTGAATGAGCATCAGGCGGCTATACCCATTGAAACCGCATGTCAG 2673
QY 157 GluLeuValAlaProGlyGluLysPheArgLeuThrIleGlyValAsnAsnGluLeuThr 176
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2674 CCGTATGTTATTCGCGGAAAGTGAATGATACATCACGTTGTGTGAACAACGAACTGAAC 2733
QY 177 HisGluThrIleProProGlyLysIleThrThrGlyAsnAlaThrGlyLysArgIleGln 196
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2734 TGGCAGACTATCCCGCGCGGAAATGGTATTAAC--GACGAAACCGCAAGAAAGACAG 2790
QY 197 ThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaArgSerIleThrPheLysSer 216
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2791 TCTTATTCATGATATTTCTTAACTATGCGGAAATCATACGACGATATGCTTACAC 2850
QY 217 ValProGlnGlnHisIleGlnAspIleThrValIleThrAspVal-----AspGlyAsp 234
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2851 AGCGCGAACAACCTGGGTGACATATCACCGGTGACGACGATGCGCGCAAGCTGTATAC 2910
QY 235 AsnGlyLeuIleAsnTyrGluValGluValAlaAsnGlnThrThrGlyGlnIleGlnIle 254
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2911 CACGCGCTGTGTGACGCGAGGAG--GTGGCCAT-----GGTATGTACAGCGTT 2958
QY 255 SerValIleAspGluAspGlyAlaIleValAlaLysAlaSerGlyValaGlnGlyThrVal 274
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2959 GAACCTCGTATGCGGATCAACAGGTGTGCACTGCAAGGACCTAGCGGACCTTTG 3018
QY 275 ThrIleProSerValLysLeuTyrGlnProGlyAlaAlaTyrLeuTyrGlnLeuGlnVal 294
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 3019 CAAGTGTGAATCCGACCTCTGGCAACGGGTGAAGTATCTATCTATGAATGTGCGTC 3078
QY 295 AsnIleValGlySerSerGlyAspValValaAspThrTyrAsnLeuAlaThrGlyValaArg 314
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 3079 -----ACAGCCAAAGCCAGACAGAGTGTATCTACCCGCTTCGCGCATCCGG 3132

```

```

QY 315 ThrValIyValAlaGlySerGlnPheLeuIleAsnGlyLysProPheThrGly 334
Db 3133 TCAGTGGACGTAAAGGCGCAAGCTTCATTAACACAAACCGTTCTTACTTACGCG 3192
QY 335 PheGlyLysHisGlyLysPheThrAlaValArgGlyLysGlyHisAspProAlaTyrMetVal 354
Db 3193 TTTGGTGGTCATGAAGTGGAGCTTAACGTGGCAAGAGATTGATTAACGTGATGATG 3252
QY 355 HisAspPheGlnLeuMetLysTyrPheGlyAlaAsnSerPheArgThrSerHisTyrPro 374
Db 3253 CACGACACGACATTAATGACATGATGGCGCAACCTCAACGCTACCTCGCATTAACCT 3312
QY 375 TyrAlaGlyLysValMetAspPheAlaAspArgAsnGlyLysValAlaIleAspGlyThr 394
Db 3313 TACGCTGAAGAGATGCTCGACGTGGCAGATGAACATGCGATCGTGGTATGATGAAC 3372
QY 395 ProAlaValGlyLeuAsnIleAlaLeu---MetGlyValSerGlyLysPro--- 412
Db 3373 GCTGCTGTGGCTTTTTCGCTCTTAAAGCATTTGTTCAAGCGGCAACAGCGGAA 3432
QY 413 GlnThrPheThrProAspAlaIleAsnAspLysThrGlnLysAlaHisLysGlnAlaIle 432
Db 3433 GAACGTACAGCGCAAGAGCGACGCAACGCGGAAACCTCAGCAACGCGCATTAACGCGCAT 3492
QY 433 ArgGlyLeuIleAlaAspLysAsnHisAlaSerValIleMetTyrSerIleAlaAsn 452
Db 3493 AAGAGCTGATGACGCGGTGACAAACACCAACCGGTGGATGGATGATGATGCGAAC 3552
QY 453 GlnProAlaSerHisGlyLysPheGlyAlaArgGlyLysPheGlyLysProLeuThrAsnLeuThr 472
Db 3553 GAACCGGATGACGCGGTGACAAAGGTGACCGGAAATATTTCGCGCACTGGCGGAAAGCAACG 3612
QY 473 ArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGlyThrAlaThrTyrGln 492
Db 3613 CGTAACTGACGACCGCGCGCTCGATCACCCTGCGTCAATGTAATGTTCTTCGACGCTCAC 3672
QY 493 LeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArgTyrPheGlyTyr 512
Db 3673 ACCGATACCTTCAGCGCTCTTGTGATGCTGTGCTGAAACCGTTATTAACGATGAT 3732
QY 513 SerGlnThrGlyAspLeuGlyLysAlaGlyAlaIleAsnGlyLysGlyLysGlyLys 532
Db 3733 GTCGAAGCGCGCATTTTGAACCGCAAGAGTACTGAAAGAACTTCGCGCTCG 3792
QY 533 GlnGlyLysPheHisArgProIleValMetThrGlyLysGlyLysAlaAspThrLeuAlaGly 552
Db 3793 CAGGAGAACTGCATCAGCGCATTAATCATCACCGAATACGCGGTGATGATGACGCGG 3852
QY 553 LeuHisSerIleLeuGlyLeuProTyrSerGlyLysPheGlnValGlnMetLeuAspMet 572
Db 3853 CTCGACTCAATGATGACCGCAATGATGATGATGATGATGATGATGATGATGATGATG 3912
QY 573 TyrHisArgValPheAspArgIleGlySerMetAlaGlyLysHisValTyrAspPheAla 592
Db 3913 TATCACCGCGCTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3972
QY 593 AspPheGlnThrAsnLeuGlyLysIleValArgValAspGlyLysGlyLysGlyLys 612
Db 3973 GATTTCGCGCTTCGCAAGCATATTCGCGGTGCGGTGCAAGAAAGGATCTTCACT 4032
QY 613 ArgAspArgLysProLysAlaAlaHisSerLeuAsnArgAlaArgTyrThrSerIleAsp 632
Db 4033 CGGACCGCAACGAAAGTGGCGCTTTTCTGCTGCAAAACGCTGACATGCGATGAC 4092

```

```

; TITLE OF INVENTION: Other Organisms
; FILE REFERENCE: 2051US
; CURRENT APPLICATION NUMBER: US/09/792,568
; CURRENT FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 11978
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: plasmid pBG363 - binary vector with the nrg2 gene
US-09-792-568-8

Alignment Scores:
Pred. No.: 2,52e-172 Length: 11978
Score: 1703.00 Matches: 336
Percent Similarity: 68.60% Conservative: 101
Best Local Similarity: 52.75% Mismatches: 178
Query Match: 50.78% Indels: 22
DB: Gaps: 11

US-10-757-093-4 (1-634) x US-09-792-568-8 (1-11978)

QY 3 PheLeuThrGlyLeuSerLeuLeuAlaAlaProSerLeuGlyThrProAlaAla 22
Db 3844 TATGTTACGTCGCGTACATTAATTAACAGTGCAGATTAATCTTACCCCATTTT 3785
QY 23 ArgHisPheProAlaArgAsnGlyMetThrGlnHisGlyLysProLeuIleLysValArgPro 42
Db 3784 TACCATCTT-----GATCGAGATTAATTAACCGGTAAATTAATTCAGTA----- 3740
QY 43 GlnArgThrSerArgGlyLysLeuValAsnLeuAspGlyLysThrLysPheAlaLeu--- 61
Db 3739 ---GAAACCCCAACCGGAAATCAAAAATCGACGCGCTGTGGCGATTCAGTCGAT 3683
QY 62 -----AlaSerGlyLeuAsnAspThrAlaGlnProThrAlaProLeuProLysGly 79
Db 3682 CGGAAACCTGCGATTAATGATCAG-----CGTTGGTGAAGAACGCGTTTCAAGAAAC 3629
QY 80 LeuGlyCysProValProAlaSerTyrAsnAspIlePheIleSerArgGlyLysHisAsp 99
Db 3628 CGGCAATGCTGTCGACAGCGAGTATTAACATGATGATGATGATGATGATGATGATG 3569
QY 100 HisValGlyTyrValTyrTyrGlnArgGlyValIleValProLysGlyTyrPheSerGln 119
Db 3568 TATCGCGGCAACCTGCTGATACCGCGAAGTCTTAATACCGAAGGTGGCGAGCGCAG 3509
QY 120 ArgTyrLeuValArgAlaGlyLysSerAlaThrHisGlyValArgTyrValAsnAsnArg 139
Db 3508 CGTATGCTGCTGCTTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3449
QY 140 LeuValAlaGlyLysValGlyLysTyrThrProPheGlyLysAlaAspValThrGlyLeuVal 159
Db 3448 GAAGTATGAGCATCAGGCGCGCTADAGCCATTTGAAGCCGATGACAGCGGTATGTT 3389
QY 160 AlaProGlyLysLysPheArgLeuThrIleGlyValAsnAsnGlyLeuThrHisGlyThr 179
Db 3388 ATTCGCGGAAAGTGTAGCATCACCGTTTGTGTAAACCAACGTAAGCTGACAGACT 3329
QY 180 IleProGlyLysLysIleThrThrGlyAsnAlaThrGlyLysArgIleGlnThrTyrGln 199
Db 3328 ATCCCGCGGAAATGATGATTAAC---GACGAAACGCGCAAGAAAGCGTCTTACTTC 3272
QY 200 HisAspPheTyrAsnTyrAlaGlyLysLeuAlaArgSerIleTyrLeuTyrSerValProGln 219
Db 3271 CATGATTTCTTAATCATGCGCGAATCCATCGACGATATGCTTACACCAACGCGGAC 3212
QY 220 GlnHisIleGlnAspIleThrValIleThrAspVal-----AspGlyAsnAsnGlyLeu 237
Db 3211 ACCTGGTGCAGCATATCAACCGGTGAGCGCATGTGCGGCAACGTAAGTCAACGCGGTCT 3152
QY 238 IleAsnTyrGlyValGlyValAlaAsnGlnThrThrGlyLysIleGlnIleSerValIle 257

```

```

Db      3151 GTTGACTGGAGG---GTGGCAAT-----GGTATGTCCAGCGTTGAAGTCCGT 3104
Qy      258 AspeluaspGlyAlaIleValAlaIleValAspSerGlyAlaGlnGlyThrValThrIlePro 277
Db      3103 GATCGGATCAACAGGTGGTGGCAACTGACACAGCACTACCGGGGCACTTTGCAAGTGGTG 3044
Qy      278 SerAllylsLeuTPGlnProGlyAlaAlaIleValIleuTyrGlnIleuGlnValAsnIleVal 297
Db      3043 AATCCGACCTCTGGCAACCGGGTGAAGGTTATCTCTATGAACCTGTGCTC-----ACA 2990
Qy      298 GlySerSerGlyAspValValAspThrTyrAsnIleuAlaThrGlyValArgThrValIlys 317
Db      2289 GCCAAAGCCAGACAGATGATATCTACCCGCTGCGCGCATCCGGTCAGTGCGCA 2930
Qy      318 ValAlaGlySerGlnPheIleuIleAsnGlyLysProPheTyrPheThrGlyPheGlyLys 337
Db      2292 GTGAAGGCGCAACGCTTCTGATTAACCAACACCGTCTTACTGCTTGGCTTGGCTGT 2870
Qy      338 HisGluAspThrAlaValArgGlyLysGlyHisAspProAlaTyrMetValHisAspPhe 357
Db      2869 CATGAAGATGGCGACTTGGCGGCAAGGATGATGATACGCTGATGTCACAGCAC 2810
Qy      358 GlnIleuMetLysTrpIleGlyAlaAsnSerPheArgThrSerHisTrpProTyrAlaGly 377
Db      2809 GCATTAATGACTGGATTTGGGGCCCACTCCTACCGTACTCGCATTAACCTTACGCTGAA 2750
Qy      378 GluValMetAspPheAlaAspArgAsnGlyIleValAlaIleAspGlnThrProAlaVal 397
Db      2749 GAGATGCTCAGCTGGCGAGATGAACATGCGATGCTGTGATGATGAATGAACTGCTGTGTC 2690
Qy      398 GlyLeuAsnIleAlaLeu---MetGlyValSerGlnSerGlyAlaPro---GlnThrPhe 415
Db      2689 GGCCTTAACTCTCTTAAAGCATTTGGTTGGAAGCGGCAACAGCGCAAGAACTGCTAC 2630
Qy      416 ThrProAspAlaIleAsnAspLysThrGlnGluAlaHisLysGlnAlaIleArgGlnLeu 435
Db      2629 AGCGAAGAGGAGTCAACGGGAACTCAGCAAGCGCATTAAGCGATTAAGAGCTG 2570
Qy      436 IleAlaArgAspLysAsnHisAlaSerValValMetTrpSerIleAlaAsnGluProAla 455
Db      2569 ATAGCCGCTGACAAAACCAACCAAGGCTGATGTGAGATTTCCCAAGCAACCGGAT 2510
Qy      456 SerHisGluAspGlyAlaArgGlyLysPheGlnProLeuThrAsnLeuThrArgGlnLeu 475
Db      2509 ACCGCTCCGACAGGTGACCGGAATTTTCGCCCACTGGCGGACCAACCGCTAAACTC 2450
Qy      476 AspProThrArgProIleThrPheAlaAsnValGlyThrAlaThrTyrGlnIleuAspArg 495
Db      2449 GACCCGACCGCTCCGATCACCTGCGTCATGTATGTTCGCGACCGTCAACCGCATCC 2390
Qy      496 IleSerAspLeuPheAspValSerCysIleAsnArgTyrPheGlyTyrPyrSerGlnThr 515
Db      2389 ATCAGAGATCTCTTGTATGTGCTGCTGCGCAACCGTTATTAACGATGATGTCGAAC 2330
Qy      516 GlyAspLeuGlnGluAlaGluAlaIleuGlnLysGlnIleuHisGlyTyrProlGlnLys 535
Db      2329 GGGGATTTGGAAACGCAAGAAAGTATCGGAAAAAACTTCTGGCCGCGCAGAGAAA 2270
Qy      536 PheHisArgProIleValMetThrGlnTyrGlyAlaAspThrIleuAlaGlyLeuHisSer 555
Db      2269 CTGCATCACCGCATTAATCAACCGAATACCGCGTGAATCTTACCGCGGCTGCATCA 2210
Qy      556 IleLeuGlyLeuProTrpSerGlnGluPheGlnValGlnMetLeuAspMetTyrHisArg 575
Db      2209 AGTACACCGACATGGAGTGAAGTATCAGTGGATGCTGATGATGATGATACCGC 2150
Qy      576 ValPheAspArgIleGlnSerMetAlaGlyLysIleValIleThrAsnPheAlaAspPheGln 595
Db      2149 GCTCTTGAATCGGATCGCGCTGCTGCTGAACAGGATGGAATTTTGGCCCATTTTTCG 2090
Qy      596 ThrAsnLeuGlyIleIleArgValAspGlyAsnLysGlyValPheThrArgAspArg 615

```

```

Db      2089 ACCTCCGACGCAATATTGGCGCTTGGCGGTACACAGAAAGGATCTTCACTGCGACCGC 2030
Qy      616 LysProLysAlaAlaHisSerLeuArgAlaArgTyrThrSerIleAsp 632
Db      2029 AAACCGAAGTGGCGGCTTTTCTGCTGCAGAAAAACGTGAGTGGCATGAAC 1979

RESULT 10
US-09-792-568-9/c
; Sequence 9, Application US/09792568
; Patent No. 669621
; GENERAL INFORMATION:
; APPLICANT: Klotz, Andreas
; APPLICANT: Davis, Keith
; TITLE OF INVENTION: No. 669621el Selectable Marker in Plants and
; FILE REFERENCE: Other Organisms
; CURRENT APPLICATION NUMBER: US/09/792,568
; CURRENT FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 12438
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Plasmid pPG361
US-09-792-568-9

Alignment Scores:
Pred. No.: 2,68e-172 Length: 12438
Score: 1703.00 Matches: 336
Percent Similarity: 68.60% Conservative: 101
Best Local Similarity: 52.75% Mismatches: 178
Query Match: 50.78% Gaps: 22
Db: 4 Indels: 11

US-10-757-093-4 (1-634) x US-09-792-568-9 (1-12438)
Qy      3 PheLeuThrGlyLysSerLeuSerLeuAlaIleProSerLeuGlyThrProAlaIle 22
Db      3844 TATGTTAGTCGCGGATCAATTAATTAACAGTGCAGATTAACTTACCGCATTTT 3785
Qy      23 ArgHisPheProArgAsnGlnMetThrGlnHisGlnGlnProLeuIleLysValArgPro 42
Db      3784 TACCATCTT-----GATCTAGATTAACGCTGTAATTAATTAATTAACAGTA----- 3740
Qy      43 GlnArgThrSerSerArgGlnLeuValAsnLeuAspGlyLeuTrpLysPheAlaLeu--- 61
Db      3739 ---GAAACCCCAACCCGGGAATCAAAAACCTGACCGCTGTGGCATTCAGTCTGGAT 3683
Qy      62 -----AlaSerGlyLeuAsnAspThrAlaGlnProTPhrAlaProLeuProLysGly 79
Db      3682 CGCGAAACTGTGAATGATCAG-----CGTTGGTGGAAACCGCTTACAAAGAAC 3629
Qy      80 LeuGlnCysProValProAlaSerTyrAsnAspIlePheIleSerArgGlnIleHisAsp 99
Db      3628 CGGCAATTGCTGTGCCAGCAGATTTAACGATCAGTTCCGCGATGACATATTCGTAAT 3569
Qy      100 HisValGlyTrpValLysTyrGlnArgGluValIleValProLysGlyTrpSerGlnGlu 119
Db      3568 TATGGGCAACGCTGGTATCAGCGCAAGTCTTATTCGGAAGGTTGGCGAGGCGCAG 3509
Qy      120 ArgTyrLeuValAlaGlyAlaGlnSerAlaThrHisHisGlyArgIleTyrValAsnAsnArg 139
Db      3508 CGTATCGTGGCTGCTTTCGATCGCGTCACTATTAACGCAAAAGTGGGCTCAATTAATCAG 3449
Qy      140 LeuValAlaGlnIleHisValIleGlyLysTyrThrProPheGlnAlaAspValIleGlnLeuVal 159
Db      3448 GAAGTGATGAGCATCAGGCGGCGCTATACCGCATTTGAACCGCATGTCAGCGCTATGTT 3389
Qy      160 AlaProGlyLysLysPheArgLeuThrIleGlyValAsnAsnGlnLeuThrHisGlnThr 179
Db      3388 ATTGCCGGAAAGTGAATGATCAACGTTTGTGTGAACAAAGAACTGAACCTGGCAGACT 3329

```

```

QY 180 ILEPROGILYLYSRLTHRGILYASNAATHRGILYLSARGILEGINTHRYGILN 199
DB 3328 ATCCGCCCGGGAATGATGATTAC---GACGAAAACGGCAGAGAAAGCAAGCTTCTTAC 3372
QY 200 HISAPSPHETRYASNTYLAAGLYLEUALARGSERILETRPHEUTYRSEVALPROGIN 219
DB 3271 CATGATTTCTTTAATCATATCCCGGAATCCATCCGAGGTATGCTTACACACGCGGAAAC 3212
QY 220 GLNHISILGLINASPILETHRVALLTHRAPVAL-----ASPGLYASPAANGLYLEU 237
DB 3211 ACCCTGGTGACGATATCACCGGTGTGACGATCGCGCAACATGTAACCAACGCGGTCT 3152
QY 238 ILEAENTYGLVALGLIVALAIAANGINTHTRGILGLINILEGINLIESERVALILE 257
DB 3151 GTTGACTGGCAGGTG---GTGGCCAAAT-----GGTATGTCACCGCTTGAACTGCGT 3104
QY 258 ASPGLUASPGLYALALALEVALALALYSAIASERGLYALAGINGLYTHRVALLTHRIEPRO 277
DB 3103 GATGCGGATCAACAGGTGGTTGCACTGGACAGACATAGCGGACTTTGCAAGTGCTG 3044
QY 278 SERVALLYLEUTTRPGLINPROGILYALALALTYRLEUTYRGLINLEUGLINVALIAENILEVAL 287
DB 3043 AATCCGACCTCTGGCAGACCGGGTGAAGGTTATCTGATGAACTGTGCGCTC-----ACA 2990
QY 298 GLYSESERGLYASPVALLAPRTHRYASNLEUALATHRGILYVALARGTHRVALLYS 317
DB 2989 GCCAAAAGCCAGACAGAGTGTGATATCTCCCGCTCGCTCGCTCAATCCGGTAGTGCA 2930
QY 318 VALALAGLYSERGLINPHEULILEANGLYLYSPROPHETRYRPHETRGILYPHEGLYLYS 337
DB 2929 GTGAAGCGGGAACAGTCTCTGATTAACCAACACGTTCTACTTCTGCTTGTGCTGCT 2870
QY 338 HISGLIAPRTHRALVALARGGLYLYSGLYNHSAPPROALATYRMEVALHISAPRPH 357
DB 2869 CATGAAGATGCGGACTTGGCTGGCAAGATTGATACCTGCTGATGGTGCAGCAC 2810
QY 358 GLNLEUMETRYETRILEGILYLAANSEPRHETRYRSETHRYRPROTYRVALAGILN 377
DB 2809 GCATTATATGACTGTGATGGGCAACTCTTACCGATCTCGATTAACCTTAAAGCTGAA 2750
QY 378 GLUVALMETAPRPHETALASPARGAENGILYLEVALIILEAPGLIUTHRPROALAVAL 397
DB 2749 GAGATCTCTGACTGGGCAATGAACATGCGATGCTGTGATTGATGAACCTGCTGCTC 2690
QY 398 GLYLEUASNILEALALEU---METGLYVALSERGLISERGLYALAPRO---GLINTHRPHE 415
DB 2689 GGCTTTAACTCTCTTAGGCAATTGGTTGCAAGCGGCAACAGCCGAAAGAACTGTAC 2630
QY 416 THRPROASPILALEASAPRYSTRHNGILUALHILEYEGILNALIILEARGILILEU 435
DB 2629 ACGGAAGAGGACTCAACGGGCACTCAAGCAAGCCACTTACAGCGATTAAGAGCTG 2570
QY 436 ILEALARGAPRYASNHISALASERVALIIMETTRPSETRILEAANGIUPROALA 455
DB 2569 ATAGCCCGTGAAGAAAACCAACCAACCGTGTGATGTGAGATTTGCCAAGCAACCGGAT 2510
QY 456 SERHISGLIAPGLYALALARGGLYRPHETGLIUPROLEUTHRANLEUTHRARGILILEU 475
DB 2509 ACCCGTCCCAAGGTGACAGCGGAATTTTCGCCCACTGGCGGAACCAACCGCTAAATC 2450
QY 476 ASPPROTHRYRPROILETHRPHETALASPVALLGLYTHRALATHRYRGLINLEUAPRARG 495
DB 2449 GACCCGACCGCTCGATCACTGCTCAATGTAATGTTCTGCGACCGCTCACACCGATACC 2390
QY 496 ILESERAPLEUPHETAPRYASERCYHILEASNARYRPHETRYRYSERGLINTHR 515
DB 2389 ATCAGGAGCTCTTTGATGTGCTGCTGCAACCGTTATTAACGATGATGCTCAAGC 2330
QY 516 GLYAPRLEUGLILUALAGLUALALALEUGLILYSGILNLEUHIISGLYTRPGLINILYLS 535
DB 2329 GGGGATTTGGAAACGCGCAGAGAGTACTGGAAAAAGAACTTCTGCGCTGGCAGAGAAA 2270

```

```

QY 536 PHEIARGPROIILEVALMETRTHRGILYLYALASPTHRLEUALAGLYLEUHIISER 555
DB 2269 CTGCATCAGCCATTTATCATCACCGGAATACCGGTGATGCTTACCGGGCTGCATCA 2210
QY 556 ILELEUGLYLEUPROTRPSERGLIUPHEGLINVALIGLIMETLEUAPMETYRHIARG 575
DB 2209 ATGTACACGACATGAGAGTGAAGATGATCATGCTGATGATGATGATGATGATGATGAT 2150
QY 576 VALPHEASPARGLILEUERMELALAGLYLHISVALITRASNPHETALASAPRPHGIN 595
DB 2149 GTCTTTGATCGGCTGACCGCGCTGCTGGTGAACAGGTATGAAATTTCCGCAATTTGCG 2090
QY 596 THRASNLEUGLYILEILEARVALAEPGLYASMLYSGLYVALPHERTHRARGAPARG 615
DB 2089 ACTGCCAAGCATATTGCGCGCTTGGCGGTGAACAAAGAGATCTTCACTCCGCAACCGC 2030
QY 616 LYSPPROLYALALALAHISERLEUARGALARGYTRPHTSERILEASP 632
DB 2029 AAACGAGTCGCGGCTTTCTGCTGCAAAAACGCTGACATGCAATGAC 1979

RESULT 11
US-09-182-117-1
; Sequence 1, Application US/09182117
; Patent No. 6204436
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Transgenic Plant
; NUMBER OF SEQUENCES: 27
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/182,117
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ. ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8012 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; US-09-182-117-1

Alignment Scores:
Pred. No.: 2 05e-171 Length: 8012
Score: 1692.00 Matches: 337
Percent Similarity: 68.35% Conservative: 95
Best Local Similarity: 53.32% Mismatches: 180
Query Match: 50.45% Indels: 20
DB: 3 Gaps: 11

US-10-757-093-4 (1-634) x US-09-182-117-1 (1-8012)
QY 10 LEUSERLEUALALAPROSERLEUGLYTRHPRIOALALARGHISAPRTHRGANGILN 29
DB 3601 CTATCTTGGCAAGACCTTCTCTATATAGAGATTCATTTCAATTTGAGAGACAG 3660
QY 30 METTHGLNHISGLINPROLEUULEYVALARGPROGLINARGHTRISERVALRGLI 49
DB 3661 CTGACAGCTGACTTACAGAGATCTCATGCTCCCTGTGAAACCCCAACCGCTGAA 3720
QY 50 LEUVALASNLEUASPGLYLEUTRPLYSRPHETALALEU-----ALASERGLYLEASN 66
DB 3721 ATCAAAAACCTGACGCGCTGTGGCATTCAGCTGATCGCAAAAACGTGAAATTGAT 3780
QY 67 ASPTRHLAGLNPRTHRALAPROLEUPROLYSGLYLEUGLILYCYSPROVALPROALA 86
DB 3781 CAG-----CGTTGTGGGAAAGCGGCTTACAAAGAACCGGGCAATTGCTGTGCCAGGC 3834

```

```

QY      87 SerTyrAsnAspIlePheIleSerArgIleHisAspHisValGlyTrpValTyrTyr 106
DB      3835 AATTTTAAGATCAGTGTCCGATGCAATATTCGAATTATTCGGCAACCTCTGAT 3894
QY      107 GlnArgGlnValIleValProlyserGlyTrpSerGlnGluArgTyrLeuValArgAlaGlu 126
DB      3895 CAGCCGGAAGCTTTATCCGAAAGGTGGCAGCGCATCGTGTGCGTTCCGAT 3954
QY      127 SerAlaThrHisHisGlyValIleTyrValAsnAspArgLeuValAlaGluHisValGly 146
DB      3955 GCGGATCACTATTCACGCAAAAGTGGTCATATACAGAAAGTATGAGATGACATGAGGC 4014
QY      147 GlyTyrThrProPheGlnValAspValThrGluLeuValAlaProGlyGluLysPheArg 166
DB      4015 GGGTATACCCCATTTGAGCCGATGTCAAGCCGTATGTTATTCGCGGAAAGGTACCT 4074
QY      167 LeuThrIleGlyValAlaAsnGluLeuThrHisGlyThrIleProProGlyLysIleThr 186
DB      4075 ATCACCGTTGTGTGAACAACGAACTGAACCTGCGACATATCCCGCGGAAATGTGAT 4134
QY      187 ThrGlyAsnAlaThrGlyLysArgIleGlnThrTyrGlnHisAspPheTyrAsnTyrAla 206
DB      4135 ACC---GACGAAACCGCAAGAAAGACGCTTACCTTCATGATTTCTTAACTATGCC 4191
QY      207 GlyLeuAlaArgSerIleTrpLeuTyrSerValProGlnGlnHisIleGlnAspIleThr 226
DB      4192 GGATCATCGTCAGCGTATGATCTTACACCGCCGAAACCTGGGTGACGATATCAC 4251
QY      227 ValValThrAspVal-----AspGlyAspAsnGlyLeuIleAsnTyrGluValGluVal 244
DB      4252 GTGTGACGATGTTCGGCAGACGTAACCAACCGCTTGTACCTGGCAGCG---GTG 4308
QY      245 AlaAsnGlnThrThrGlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleVal 264
DB      4309 GCCAAT-----GGTGAATGTCAGCGTTGAACCTGCGATGCGATCAACGATGCGTT 4359
QY      265 AlaLeuAlaSerGlyAlaGlnGlyThrValThrIleProSerValLysLeuTrpGlnPro 284
DB      4360 GCAACTGGAACAGGCACTACCGGACCTTTCAGATGTAATCGCACCTCTGCAACCG 4419
QY      285 GlyAlaAlaIleTyrLeuTyrGlnLeuGlnValAsnIleValGlySerSerGlyAspValVal 304
DB      4420 GGTAAGATTTATCTCTATGAACCTGTGGTC-----ACAGCAAAAGCCAGACGATGT 4473
QY      305 AspThrTyrAsnLeuAlaThrGlyValArgThrValLysValAlaGlySerGlnPheLeu 324
DB      4474 GATATCTACCGCTTCGCGTCGCGCATCGGTCACTGCGAGTGAAGGCGCAACGATTCCTG 4533
QY      325 IleAsnGlyLysProPheTyrPheThrGlyPheGlyLysHisGluAspThrAlaValArg 344
DB      4534 ATTACCAACAAACCGTTTACTTACTGAGCTTTGGTGTGCTCAAGATGCGGACTTACGT 4593
QY      345 GlyLysGlyHisAspProAlaTyrMetValHisAspPheGlnLeuMetLysTrpIleGly 364
DB      4594 GCGAAGATTCGATTAACGTCTGATGTGCAACCAACGATTAATGACTGATGGG 4653
QY      365 AlaAsnSerPheArgThrSerHisTyrProTyrAlaGluGluValMetAspPheAlaAsp 384
DB      4654 GCCAATCTCACTGATCTCCCATTAACCTTAACGTGAAGAGATGCTGACTGGGAGAT 4713
QY      385 ArgAsnGlyIleValIleAspGluThrProAlaValGlyLeuAsnIleAlaLeu--- 403
DB      4714 GAACATGGCATCTGATGATGATGAACCTGCTGTGCGCTTAACTCTTTAGGC 4773
QY      404 MetGlyValSerGluSerGlyAlaPro---GlnThrPheThrProAspAlaIleAsnAsp 422
DB      4774 ATTGGTTTCGAAGCGCGCAACGCAAGCAAAAGACTGTACAGCAAGAGCGATCAACCGG 4833
QY      423 LysThrGlnGlnAlaHisIleGlyGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHis 442
DB      4834 GAAACTCAGCAAGCGCACTTACAGCGATTAAGAGCTGATAGCGCGTGAACAAACACAC 4893

```

```

QY      443 AlaSerValValMetTrpSerIleAlaAsnGluProAlaSerHisGluAspGlyValArg 462
DB      4894 CCAAGCTGTGATGTGAGATGATGCCAAGCAACCGGATACCGGTCT-----GCACGG 4947
QY      463 GluTyrPhe-----GluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgPro 480
DB      4948 GAATATTTCCGCAATTGCGCACCTGCGGAAAGCAACGGGTAACTCGACCGGACGCTCCG 5007
QY      481 IleThrPheAlaAsnValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPhe 500
DB      5008 ATCACCTGCGTCAATGATATGTTCTGCAACCTCAACCAACATACATCAGGATCTCTTT 5067
QY      501 AspValSerCysIleAsnArgTyrPheGlyTrpTyrSerGlnThrGlyAspLeuGluGlu 520
DB      5068 GATGCGCTGCTCGTGAACCGTTATTAACGATGATGATCCAAACGCGCGATTTGGAAACG 5127
QY      521 AlaGluAlaAlaLeuGluLysGluLeuHisGlyTyrGlnGluLysPheHisArgProIle 540
DB      5128 GCAGAGAAAGTACTGGAAGAAAGAACTTCTGCGCTGGCAGGAAATGTCATCAGCGAT 5187
QY      541 ValMetThrGluTyrGlyValAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuPro 560
DB      5188 ATCATCACCGAATACGCGGTGATACCTTACCGCGGTGCACTCAATGTACACGACATG 5247
QY      561 TrpSerGluGluPheGlnValGlnMetLeuAspMetTyrHisArgValPheAspArgIle 580
DB      5248 TCGAGTGAAGAGTATCAGTGTGATGCTGCGATGTATGTATCAACCGGTGTTGATGCGCTC 5307
QY      581 GluSerMetAlaGlyGluHisValThrAsnPheAlaAspPheGlnThrAsnLeuGlyIle 600
DB      5308 ACGGCGTGTGCGTGAACAGGTATGAATTTCCCGGATTTTGGCACTCCGAAAGGCTTA 5367
QY      601 IleArgValAspGlyAsnLysLeuGlyValPheThrArgAspArgLysProLysAlaAla 620
DB      5368 TTGCGCGTTGGCGGTAAACAGAAAGGATCTTCACTCGGACCGCAACCGAAGTCGCG 5427
QY      621 AlaHisSerLeuArgAlaArgTrpThrSerIleAsp 632
DB      5428 GCTTTCTCTGTCGCAAAACGCTGGATGCGCATGAAC 5463

```

```

RESULT 12
US-09-434-039A-1
; Sequence 1, Application US/09434039A
; Patent No. 6531649
; GENERAL INFORMATION:
; APPLICANT: MANNERLOEF, Marie
; APPLICANT: TENNING, Paul Peter
; TITLE OF INVENTION: Transgenic Plants
; FILE REFERENCE: 09/434,039
; CURRENT APPLICATION NUMBER: US/09/434,039A
; PRIOR FILING DATE: 1999-11-04
; PRIOR APPLICATION NUMBER: 60/112,003
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 09/182,117
; PRIOR FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 8012
; TYPE: DNA
; ORGANISM: Sugar beet
US-09-434-039A-1

```

Alignment Scores:

```

Pred. No.: 2,05e-171 Length: 8012
Score: 1692.00 Matches: 337
Percent Similarity: 68.35% Conservative: 95
Best Local Similarity: 53.32% Mismatches: 180
Query Match: 50.45% Indels: 20
DB: 4 Gaps: 11

```

US-10-757-093-4 (1-634) x US-09-434-039A-1 (1-8012)

QY	10	LeuSerLeuAlaIaIaProSerLeuGlyThrProAlaAlaArgHisLeuProArgAspGlu	29
Db	3601	CTATTCCTTCGCAAGACGCTTCCTCTCTATATGAAGAACTTCATTCTTGGAGAGACACG	366C
QY	30	MetThrGlnHisGluGlnProLeuHisLeuValArgProGlnArgThrSerSerArgGlu	49
Db	3661	CTGACAGAGCTGACTCTAGCAAGCTTCATCCATGGTCGGTCTGTGAAGAAACCCCAACCGGTGA	3720C
QY	50	LeuValAsnLeuAspArgGlyLeuThrIleuPheLeuAlaLeu-----AlaSerGlyLeuAsn	66
Db	3721	ATCAAAAACCTCGACGGCTGTGGCATTCAGTCTGGATTCGCGAAACCTGTGGAAATTGAT	3780C
QY	67	AspThrAlaGlnProTyrThrAlaProLeuProGlyLeuGluCysProValProAla	86
Db	3781	CAG-----CGTTGGTGGGAAGAGCCGCTTACAGAAACCCGGCAATTCGTCTGCGACGC	3834C
QY	87	SerTyrAsnAspIlePheIleSerArgGluIleHisAspHisValGlyTyrValTyrTyr	106
Db	3835	AGTTTAAACGATTCAGTTGCGCGATCGAGATTTCTGTAATTTATGCGCGCAACGTCTCGCAT	3894C
QY	107	GlnArgGluValIleValProIleGlyTyrPheSerGlnGluArgTyrLeuValArgAlaGlu	126
Db	3895	CAGCGCAAGCTCTTATATACGAAAGCTTGGGAGGCCAGCGTATCTGTCTGGTTTCAT	3955C
QY	127	SerAlaThrHisIleGlyArgIleTyrValAsnAspArgLeuValAlaGluHisValGly	146
Db	3955	CGCGTCACCTCATTAACGGAAAGTGGGTCAATATACAGAAAGTATGATGACATCAGGCGC	4014C
QY	147	GlyTyrThrProPheGlnAlaAspValThrGluLeuValAlaProGlyGluIlePheArg	166
Db	4015	GGCTAATAGCCATTGTAAGCCGATGTCCAGCGGTAATGTATTTATTCGGGAAAAAGTGTAGT	4074C
QY	167	LeuThrIleGlyValAsnAsnGluLeuThrHisGluThrIleProProGlyIleIleThr	186
Db	4075	ATCACCCGTTTGTTGTGAACAAGACTGAATCTGGCAGACTATTCGCCCGGAAATGGTATT	4134C
QY	187	ThrGlyAsnAlaThrGlyIleYsaArgIleGlnThrTyrGlnHisAspPheTyrAsnTyrAla	206
Db	4135	ACC---GACGAAAAACGGCAAAAAAGACAGCTTACTTCATGATTTCTTTAACTATGCC	4191C
QY	207	GlyLeuAlaArgSerIleThrPheTyrSerValProGlnGlnHisIleGlnAspIleThr	226
Db	4192	GGAAATTCATCCGACAGGTAAAGTCTCTACACACGCCGAACCTGGGTGACGATATCCAC	4251C
QY	227	ValValThrAspVal-----AspIleYsaAsnGlyLeuIleAsnTyrGluValGluVal	244
Db	4252	GTGGTGAAGCATGTGCGCAAGACTGTAAACCAACGCGGTGTGTTGACTGCAAGTGT---GTG	4308C
QY	245	AlaAsnGlnThrThrGlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleVal	264
Db	4309	GCCAAAT-----GGTAGTGTACGCGCTTAACCTGATGCGGATCAACAGGTGGTT	4359C
QY	265	AlaIleYsaIaSerGlyAlaGlnGlyThrValThrIleProSerValIleLeuThrPheGlnPro	284
Db	4360	GCAACTGCAACAAGCAGCACTAGCGGAGACTTTCGAAGTGTGAATCCGCACTCTGCGAACCG	4419C
QY	285	GlyAlaAlaTyrLeuTyrGlnLeuGlnValAsnIleValIleGlySerSerGlyAspValVal	304
Db	4420	GGTAAAGGTTATCTCTATGAAGTGTGCGTCTC-----ACAGCCAAACCGCAACAGAGTGT	4473C
QY	305	AspThrTyrAsnLeuAlaIaThrGlyValArgThrValIleYsaIleAlaGlySerGlnPheLeu	324
Db	4474	GATATCTAACCCGCTTCGCGTGGCATCCGGGTGACGTGAGCGAGTGAAGGGCAACAATTCTTG	4533C
QY	325	IleAsnGlyIleProPheTyrPheThrGlyPheGlyIleYsaIleGluAspThrAlaValArg	344
Db	4534	ATTAACCAACAACCGTTCTACTTACTGGCTTGGTGTGTCATGAAGATGCGGACTTACGT	4593C
QY	345	GlyIleGlySerGlnHisAspProAlaTyrIleMetValHisAspPheGlnLeuMetIleTyrIleGly	364
Db	4594	GGCAAAAGGATTCGATTAACCTGTGATGGTGGACGACCAACGATTAATGAGCTGGAATTTGGG	4653C

```

Oy      365  AAlaMSerPheAqgThSerSniStyPProTylAaGluGluValMeCArPheAlaAsp  384
Db      4554  GCCAACTCCrAcCTAcCTGCcATTACcCTTAcCGcTAAGAGAcGTCCGAATGGCGcGAT  4713
Oy      385  ArgAnGlyIleValIleAspGluThrProAlaValGlyLeuAnIleAlaLeu---  403
Db      4714  GAACATGCGCATCGTGGATGATGAAGAACTGCTCTGCGGCTTTAACTCTCTTAAAGC  4773
Oy      404  MetGlyValSerGluSerGlyAlaPro---GlnThrPheThrProAlaPheAlaAsp  422
Db      4774  ATTGGTTTCGAGCGGGCAACAAcCCCAAGAACTGTACAGCGAAGCGcAGTCAcGGG  4833
Oy      423  LysThrGlnGluAlaHisIlySglAlaIleArgGluLeuIleAlaAspIlyAsnHis  442
Db      4834  GAACCTCAGCAAGCGCACTTACAGCGCATTTAAAGcGTAGTAgCGCGTACAAAAACAC  4893
Oy      443  AlaSerValAlaMetTPSerIleAlaAsnGluProAlaSerHisGluAspGlyAlaArg  462
Db      4894  CCAAGCGTGGTGAATGGATGGAGTATTGGCCAGAAcCGGATACCCGCTCT-----GCAcGG  4947
Oy      463  GluTyPhe-----GluProLeuThrAsnLeuThrArgGluLeuAspProThrArgPro  480
Db      4948  GAATATTTCGcCATTTGGcCATCGCACTGGCGGAGCAACCGGTAAACTCGAACCGCACTCGCG  5007
Oy      481  IleThrPheAlaAsnValGlyThrAlaIleThrArgGluLeuAspArgIleSerAspLeuPhe  500
Db      5008  ATCAcCGTGGCTCAATGTAAATGTTCTGGCAcCGCTCACCCATATCAATAGCGATCTCTTT  5067
Oy      501  AspValSerCysIleAsnArgTyPheGlyTyPTrPzSerGlnThrGlyAspLeuGluGlu  520
Db      5068  GATTCGCTGTCcCGCAACCGGTATTACGATGGTATGCCAAAGCGCGCGATTGGAAAGC  5127
Oy      521  AlaGluAlaAlaLeuGluTyPLeuHisGlyTyPTrGlnGluTyPLeuHisArgProIle  540
Db      5128  GCAGAGAAGGTACTGAAAAAAGAACTTCTGGCGCTGGCAGAGAAACTGCATCAGCCGATT  5187
Oy      541  ValMetThrGluTyPArgIlyAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuPro  560
Db      5188  ATCTATCAcCGAAATCGCGCGTGGATACGTTAGCCGGCTGCATCAATGTACACCGACATG  5247
Oy      561  TrpSerGluGluPheGlnValAlaGlnMetLeuAspMetTyTrHisArgValPheAspArgIle  580
Db      5248  TGGATGTAAAGGTATCAAGTGTGCATGGCTGGATATGTATCAcCGCGCTTTGTATCGCGTC  5307
Oy      581  GluSerMetAlaGlyGluHisValIleTPAsnPheAlaAspPheGlnThrAsnLeuGlyIle  600
Db      5308  AGCCCGCTCGTGGCTGAACAGGTATGGAATTCGCCGATTTTGGCACTCGCAAGGcGATA  5367
Oy      601  IleArgValAspGlyAsnTyPLeuGlyValPheThrArgAspArgIlyProIlyValAla  620
Db      5368  TTGGCGCTTGGCGGTTACAAGAAAGGAGATCTTCACTCGCAcCGCAACCGAAGTCCGGCG  5427
Oy      621  AlaHisSerLeuArgAlaArgTyPThrSerIleAsp  632
Db      5428  GCTTTCTCTGCAAAAACCGCTGAGCTGCATGAC  5463

RESULT 13
US-09-182-117-5
; Sequence 5, Application US/09182117
; Patent No. 620436
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Transgenic Plant
; NUMBER OF SEQUENCES: 27
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/182.117
; FILING DATE:
; CLASSIFICATION:

```

```

; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8418 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-09-182-117-5

Alignment Scores:
Pred. No.: 2,21e-171 Length: 8418
Score: 1692.00 Matches: 337
Percent Similarity: 68.35% Conservative: 95
Best Local Similarity: 53.32% Mismatches: 180
Query Match: 50.45% Indels: 20
DB: Gaps: 11

US-10-757-093-4 (1-634) x US-09-182-117-5 (1-8418)
QY 10 LeuSerLeuAlaIaProSerLeuGlyThrProAlaAlaArgHisPheProArgAenglu 29
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3580 CTAATCTTCGCAAGACCCCTTCCTATATAGGAAGTTCATTGAGAGACAG 3639
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 30 MetThrGlnHisGluGlnProLeuIleValArgProGlnArgThrSerSerArgGlu 49
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3640 CTGACAAAGCTGACTTACAGATCTCCATGCTCCGCTCTGTAGAAACCCCAACCCGTGAA 3699
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 50 LeuValAsnLeuAspGlyLeuThrPlysPheAlaLeu-----AlaSerGlyLeuAsn 66
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3700 ATCAAAATTCAGACGGCTGTGGGCACTTCAGTCTGCATGCCAAACCTGGAAATTGAT 3759
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 67 AspThrAlaGlnProThrAlaProLeuProLysGlyLeuGlyCysProValProAla 86
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3760 CAG-----CGTGTGGGAAAGCGCGCTTACAAGAAACCGGCAATTGCTGTGCCAGGC 3813
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 87 SerThrAsnAspIlePheIleSerArgGlnIleHisAspHisValGlyThrValTyrTyr 106
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3814 AGTTTACCATCATGTCGCCGATGACGATATTCGATTAATGCGGCAACGCTGGTAT 3873
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 107 GlnArgGluValIleValProLysGlyTyrSerGlnGluArgTyrLeuValArgAlaGlu 126
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3874 CAGGCGAAGCTTTATACCGAAAGTTGGGCAAGCGGTATCGTGCCTTCGAT 3933
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 127 SerAlaThrHisHisGlyArgGlyLeuValAsnAsnArgLeuValAlaGlnHisValGly 146
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3934 CGCGTCACTCATTCGCGCAAGTGTGGTCAATATCAGAGAGATGATGAGATCAGGCG 3993
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 147 GlyTyrThrProPheGlnIleAspValThrGluLeuValAlaProGlyGluLysPheArg 166
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3994 GCGTATACGCGATTTGAAGCGATGTCACCGCATGTATTATGCCGGAAGATGATCCT 4053
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 167 LeuThrIleGlyValAsnAsnGluLeuThrHisGluThrIleProProGlyLysIleThr 186
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4054 ATCAACCGTTTGTGTGAACAACTGAATCGACACTATCCCGCGGAGATGTGAT 4113
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 187 ThrGlnAsnAlaThrGlyLysArgGlyLeuGlnThrTyrGlnHisAspPheTyrAsnTyrAla 206
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4114 ACC---GACGAAACCGGCAAGAAAGCGATCTTCACTTCATGATTTTAACTATGACC 4170
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 207 GlyLeuAlaArgSerIleThrPheLeuTyrSerValProGlnGlnHisIleGlnAspIleThr 226
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4171 GGAATCCATCGGAGCGTATGCTCTACACCAAGCGCAACCTGGGTGAGATATACACC 4230
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 227 ValValThrAspVal-----AspGlyAspAsnGlyLeuIleAsnTyrGlnValGluVal 244
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4231 GTGGAGAGCATGTGCGCAAGACTGTAAACACCGGCTGTGACTGCAAGGTG---GTG 4287
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 245 AlaAsnGlnThrThrGlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleVal 264
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4288 GCCAAT-----GGTGAATGTACGCGTGAATGCTGCTGATGCGGATCAACAGGTGATT 4338
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

QY 265 AlaValAspSerGlyValIleGlnGlyThrValThrIleProSerValIleLeuThrPro 284
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4339 GCAATGAGACAGGACTAGCGGAGCTTTCGATGATGAAATCCGACACTCTGGCAACCG 4398
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 285 GlyAlaAlaTyrLeuTyrGlnLeuGlnValAsnIleValGlySerSerGlyAspValVal 304
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4399 GGTGAAGGTATCTCTATGAACTGTGGCTC-----ACAGCCAAACCCAGACAGACTGT 4452
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 305 AspThrTyrAsnLeuAlaThrGlyValArgThrValIleValAlaGlySerGlnPheLeu 324
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4453 GATATCTACCCGCTTCGCGTGGCATCCGATCGATGCGAGAGAGGAGGAGCAATTCCTG 4512
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 325 IleAsnGlyLysProPheTyrPheThrGlyPheGlyLysHisGluAspThrAlaValArg 344
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4513 ATTACCAACCAACCGTTCTTACTTACCTGCTTGTGCTCATGAAAGATCGCACTAGCT 4572
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 345 GlyLysGlyHisAspProAlaTyrMetValHisAspPheGlnLeuMetLysThrIleGly 364
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4573 GGCAAAGATTCGATTAACGTGCTGATGTCAGACACACGCAATTAAATGCACTGGATTGGG 4632
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 365 AlaAsnSerPheArgThrSerHisTyrProTyrAlaGluGluValMetAspPheAlaAsp 384
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4633 GCAACTCTTACCTGATCTCGCATTCCTTACCTGATGAGAGATGCTGATGCGGACAT 4692
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 385 ArgAsnGlyIleValIleAspGluThrProAlaValGlyLeuAsnIleAlaLeu--- 403
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4693 GAAATGCGATGCTGTGATGATGATAAATCTGCTGCGCTTAACTCTCTTTAGGC 4752
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 404 MetGlyValSerGluSerGlyAlaPro---GlnThrPheThrProAspAlaIleAsnAsp 422
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4753 ATTGATTTCGAAGCGGCAACGCAAGAAAGCTGTACAGCAAGAGACACTCAACGG 4812
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 422 LysThrGlnGluAlaHisLysGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHis 442
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4813 GAACTCAGACAGCCGACTTACAGCGATTAAGAGCTGATGACCGCTGACAAACACAC 4872
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 443 AlaSerValMetTyrSerIleAlaAsnGluProAlaSerHisGluAspGlyAlaArg 462
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4873 CCAAGGTGTGATGATGGAGTATTCACCAACCGGATACCCGCTC-----CACGG 4926
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 463 GlyTyrPhe-----GluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgPro 480
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4927 GAATATTTGGCATTTCCGCACTGGCGGAAGCAACCGTAATCGAACCCGACGGCTCG 4986
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 481 IleThrPheAlaAsnValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPhe 500
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4987 ATCACTGGGTGAATGTATGTTCTGCGACGCTCACACCGATACATCAGCATCTCTT 5046
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 501 AspValSerCysIleAsnArgTyrPheGlyTyrSerGlnThrGlyAspLeuGluGlu 520
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5047 GATGTCTGTGCTGTAACCGTATTAACGATGGATGTCCAAAGCGGCAATTTGAAAGC 5106
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 521 AlaGluAlaAlaLeuGlnLysGluLeuHisGlyTyrGlnGluLysPheHisArgProIle 540
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5107 GCAAGAAAGTACTGAAAGAAAGAACTTCGGCTGCGCAAGAAATCGATCAGCGCAT 5166
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 541 ValMetThrGluTyrGlyAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuPro 560
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5167 ATCATACCGAATACGCGGTGATGATGATGACCGGCTGCACTCAATGTACACCGACATG 5226
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 561 TyrSerGlnGluPheGlnValGlnMetLeuAspMetTyrHisArgValPheAspArgIle 580
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5227 TGAAGTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5286
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 581 GluSerMetAlaGlyGlnHisValTyrAsnPheAlaAspPheGlnThrAsnLeuGlyIle 600
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5287 AGCGCCCTCGTGGTGAACAGTATGAAATTTGCGCATTTTGCACCTCGCAAGCATTA 5346
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 601 IleArgValAspGlyAsnLysLysGlyValPheThrArgAspArgLysProLysAlaAla 620
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5347 TTGCGCTGTGCGGTAAACAAAGAGATCTTACTCGCAACGCAACCGAAGTGGCG 5406
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 621 AlaHisSerLeuArgAlaArgTyrThrSerIleAsp 632
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db      5407 GCTTTCTGCTGAAAGCGTGCATGCGCATGAC 5442
RESULT 14
US-09-434-039A-5
; Sequence 5, Application US/09434039A
; Patent No. 6531649
; GENERAL INFORMATION:
; APPLICANT: MANNERLOF, Marie
; APPLICANT: TENNING, Paul Peter
; APPLICANT: STEEN, Per
; TITLE OF INVENTION: Transgenic Plants
; FILE REFERENCE: 09/434, 039
; CURRENT APPLICATION NUMBER: US/09/434, 039A
; CURRENT FILING DATE: 1999-11-04
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 09/182,117
; PRIOR FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 8418
; TYPE: DNA
; ORGANISM: Sugar beet
US-09-434-039A-5

Alignment Scores:
Pred. No.:      2,21e-171      Length:      8418
Score:          1692.00        Matches:      337
Percent Similarity: 68.35%     Conservative: 95
Best Local Similarity: 53.32%  Mismatches:    180
Query Match:    50.45%        Indels:       20
DB:              4            Gaps:           11

US-10-757-093-4 (1-634) x US-09-434-039A-5 (1-8418)
QY      10 LeuSerLeuAlaAlaProSerLeuGlyThrProAlaAlaArgHisPheProArgAsnGlu 29
Db      3580 CTATCTCTTCGCAAGACCCCTTCTCTATATPAGAGAGTTCATTTCATTTGGAGAGACAGC 3639
QY      30 MetThrGlnHisGluGlnProLeuIleValArgProGlnArgThrSerSerArgGlu 49
Db      3640 CTGACAGAGCTGACTAGCAGATCTCATGTCGCTCTGTAAAGCCCAACCCGCGAA 3699
QY      50 LeuValAsnLeuAspGlyLeuThrPheAlaLeu-----AlaSerGlyLeuAsn 66
Db      3700 ATCAAAAACCTCGACCGCGCTGTGGCATTCAGTCTGGATCGCGAAAACCTGTGAATTGAT 3759
QY      67 AspThrAlaGlnProThrPheAlaProLeuProGlyGluGlnCysProValProAla 86
Db      3760 CAG-----CGTTGGTGGAAAGCGGCTTACAAAGAACCGCGCAATTGCTGTCCAGGC 3813
QY      87 SerTyrAsnAspIlePheIleSerArgGluIleAspHisValGlyTyrValTyr 106
Db      3814 AGTTTAAAGCATAGTTCGCCGATGCGAGATTCGTAATTACGGGCAACGCTGGTAT 3873
QY      107 GlnArgGluValIleValProGlyGlyTyrSerGlnGluArgTyrLeuValArgAlaGlu 126
Db      3874 CAGCGGAGAGTCTTTATCCGAAAGGTGGCAGCGCAGCGATTCGTCGCGTTTCGAT 3933
QY      127 SerAlaThrHisIleGlyArgIleTyrValAsnAspArgLeuValAlaGlnHisValGly 146
Db      3934 GCGGTCACATTAATGCGCAAAAGTGGGTGCAATTAATCGAAGATGATGCGATCGACGGC 3993
QY      147 GlyTyrThrProPheGlnAlaAspValThrGlnLeuValAlaProGlyGlnCysPheArg 166
Db      3994 GGGTAAACCGCATTTGAAGCCGATGTCAACCCGATGTTATTCGCGGAAAGATGATCGT 4053
QY      167 LeuThrIleGlyValLeuAsnGlyLeuThrHisGluThrIleProProGlyValIleThr 186
Db      4054 ATCACCAGTTTGTGTGAACAACGAACTGACCTGCAGACTATCCCGCGGGAATGTGATTT 4113

```

```

QY      187 ThrGlyAsnAlaThrGlyLysArgIleGlnThrTyrGlnHisAspPheTyrAsnTyrAla 206
Db      4114 ACC---GACGAAGACGCAAGAAAGCATGCTTACTTCATGATTTCTTTAACTATGCC 4170
QY      207 GlyLeuAlaArgSerIleThrPheTyrSerValProGlnGlnHisIleGlnAspIleThr 226
Db      4171 GGAATCCATCGCAGCGTAATGCTCTACACACACCGCAACACTGGGGGAGCATATCACCC 4230
QY      227 ValValThrAspVal-----AspGlyAspAsnGlyLeuIleAsnTyrGluValGluVal 244
Db      4231 GGTGTACCGATGTCCGCAAGCTGTACACCGCTGTTCATCTGCGCAGGTG---GTG 4287
QY      245 AlaAsnGlnThrThrGlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleVal 264
Db      4288 GCCAAT-----GTGATGTACACGCTTGAATCTCGATTCGGATCAACAGGTGTT 4338
QY      265 AlaLeuAlaSerGlyValaGlnGlyThrValThrIleProSerValIleValLeuTyrGlnPro 284
Db      4339 GCAACTGCAACAGCGACTAGCGGACTTTCGAAGTGTGATTCGCACTCTGCGCAACCG 4398
QY      285 GlyAlaAlaTyrLeuTyrGlnLeuGlnValAsnIleValGlySerSerGlyAspValVal 304
Db      4399 GTGAAAGTTATCTATGAACTGTGCGTC-----ACAGCCAAAAGCCAGACAGTGT 4452
QY      305 AspThrTyrAsnLeuAlaThrGlyValArgThrValIleValAlaGlySerGlnPheLeu 324
Db      4453 GATATCTACCCGCTTCCGCTCGGCATCCGCTGACTGGCAAGGCGCAACGTTCTCTG 4512
QY      325 IleAsnGlyLysProPheTyrPheThrGlyPheGlyValHisGluAspThrAlaValArg 344
Db      4513 ATTAACACAAACCGTTTACTTACTTACGCTTGGTGGTGCATGAAATGCGCACTTAAGT 4572
QY      345 GlyLysGlyHisAspProAlaTyrMetValHisAspPheGlnLeuMetLysTyrIleGly 364
Db      4573 GGCAAAGGATTCATTAACGCTGATGTGACAGCAGCAGCATTAATGACGTGATTTGGG 4632
QY      365 AlaAsnSerPheArgThrSerHisTyrProTyrAlaGluGluValMetAspPheAlaAsp 384
Db      4633 GCCAATCTCTACCGTACCTTCGCTTACCTTACCGTGAAGAGATTCGTCGACTGGCAGAT 4692
QY      385 ArgAsnGlyIleValValIleAspGlyThrProAlaValGlyLeuAsnIleAlaLeu--- 403
Db      4693 GAACATGCGACTCGTGTGATGTGAACTGCTGCTGCGCTTTAACTCTCTTTAGGC 4752
QY      404 MetGlyValSerGluSerGlyAlaPro---GlnThrPheThrProAspAlaIleAspAsp 422
Db      4753 ATTGGTTTGGAAAGCGGCAACAGCGGAAGAACTGTACAGCAAGAGAGCATCAACGGG 4812
QY      423 LysThrGlnGluAlaHisLysGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHis 442
Db      4813 GAACTCAGCAAGCGCATTTACAGGCGATTAAAGACTGATACCGCGTGCACAAAACAC 4872
QY      443 AlaSerValValMetTyrSerIleAlaAsnGluProAlaSerHisGluAspGlyAlaArg 462
Db      4873 CCAAGCGGTGTATGTGAGTATTCGCAACGAAACCGGATACCGTCTCT-----GCACGG 4926
QY      463 GlyTyrPhe-----GluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgPro 480
Db      4927 GAATATTTGGCATTTTCGCCATCGCGCAAGCAACCGTAATCTCAACCGCGCGCGTCCG 4986
QY      481 IleThrPheAlaAsnValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPhe 500
Db      4987 ATCACTGCGCTCATTTATGTTCTGCGAGCTCAGACCGATACATCAGCGATCTCTTT 5046
QY      501 AspValSerCysIleAsnArgTyrPheGlyTyrTyrSerGlnThrGlyAspLeuGlnGlu 520
Db      5047 GATGTCTGTGCTGAACCGTTATTAAGATGTATGTCCAAAGCGCGCATTTGCAACG 5106
QY      521 AlaGluAlaAlaLeuGlnLysGlyLeuHisGlyTyrGlnGlnLysPheHisAspProIle 540
Db      5107 GCAGAGAGGTATCTGAAAGAACTTGTGCGCTGCGAGGAAAGTGTGACGCGCATTT 5166
QY      541 ValMetThrGluTyrGlyAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuPro 560

```

```

Db      5167 ATCATCACCAATACGCGCTGATACGTTAGCCGGCTGCATCTCAATGTACACCGACATG 5226
Qy      561 TTPSerTlnuPheglnValGlnMetLeuAspMetTyrHisArgValPheAspArgTle 580
Db      5227 TGGAGTAAAGATATCGATGCTGATGCTGATATGTATACCCGCTTTGATCGCGTC 5286
Qy      581 GluSerMetAlaGlyGlnHisValTTPAspPheAlaAspPheGlnThrAsnLeuGlyIle 600
Db      5287 AGGCGCGCTCGCTGTAACAGTATGGAATTTGCCCATTTTGGCACTTGCAGAGCATTA 5346
Qy      601 IleArgValAspGlyAsnLysLysGlyValPheThrArgAspArgLysProLysAlaIle 620
Db      5347 TTGGCGCTTGGCGGTAAACAAGAGGATCTTCATCGACCGCAACCGAAGTGGCG 5406
Qy      621 AlaHisSerLeuArgAlaArgTTPThrSerIleAsp 632
Db      5407 GCTTTTGTCTGCAAAAACGCTGACTGGCATGAC 5442

```

```

RESULT 15
US-09-182-117-4
; Sequence 4, Application US/09182117
; Patent No. 6204436
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Transgenic Plant
; NUMBER OF SEQUENCES: 27
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/182,117
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8798 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; US-09-182-117-4

```

```

Alignment Scores:
Pred. No.: 2,386-171 Length: 8798
Score: 1692.00 Matches: 337
Percent Similarity: 68.35% Conservative: 95
Best Local Similarity: 53.32% Mismatches: 180
Query Match: 50.45% Indels: 20
DB: 3 Gaps: 11

```

US-10-757-093-4 (1-634) x US-09-182-117-4 (1-8798)

```

Qy      10 LeuSerLeuAlaAlaProSerLeuGlyThrProAlaAlaArgHisAspPheArgAsnGln 29
Db      3690 CTATCTCTGGCAAGACCTTCTCTATATAAGAAAGTTCATTTCATTGAGAGGACACG 3749
Qy      30 MetThrGlnHisGlnGlnProLeuLysLysValArgProGlnArgThrSerSerArgGln 49
Db      3750 CTGACAAAGTGAAGTCTAGACAGATCTCCATGTCCTGTAAGAAACCCCAACCCGTGA 3809
Qy      50 LeuValAsnLeuAspGlyLeuTTPAspPheAlaLeu-----AlaSerGlyLeuAsn 66
Db      3810 ATCAAAAACCTGACGCGCTGTGGGATTCAGCTGGATCGCAAAACCTGTGAATGTAGT 3869
Qy      67 AspThrAlaGlnProTTPThrAlaProLeuProLysGlyLeuGlnLysProValProAla 86
Db      3870 CAG-----CGTTGGTGGGAAAGCGGCTTACAAAGAAACCGGCAATTGCTGTGGACGC 3923

```

```

Qy      87 SerTyrAsnAspIlePheIleSerArgGlnIleHisAspHisValGlyThrValTyrTyr 106
Db      3924 AGTTTAAAGATAGTTCGCCCATGCAAGATATCTGTAATTAAGCGGCAACGCTGGAT 3983
Qy      107 GlnArgGlnValIleValProLysGlyTTPSerGlnLysArgTyrLeuValArgAlaGln 126
Db      3984 CAGCGCAAGGCTTTTATACCGAAAGGTTGGGACGCGCACGATGCTGCTGGCTTCAT 4043
Qy      127 SerAlaThrHisHisGlyArgLysLysTyrValAsnAspArgLeuValAlaGlnHisValGly 146
Db      4044 GCGGTACATCATTAATCGCAAGATGGGTCAATATACAGAAAGTGTATGAGATCGAGGC 4103
Qy      147 GlyTyrThrProPheGlnAlaAspValThrGlnLeuValAlaProGlyGlnLysPheArg 166
Db      4104 GGTATACCCCATTTTAAACCCATGTACAGCCGCTATGTTATGGCGGAAAGTACCT 4163
Qy      167 LeuThrIleGlyValAsnAsnGlnLeuThrHisGlnThrIleProProGlyLysIleThr 186
Db      4164 ATCACCGTTTGTGTGAACAACGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 4223
Qy      187 ThrGlyAsnAlaThrGlyValArgLysGlnThrTyrGlnHisAspPheTyrAsnTyrAla 206
Db      4224 ACC--GACGAAACCGCAAGAAAGACGCTTCTTCATGATTTCTTTAATCTATGCC 4280
Qy      207 GlyLeuAlaArgSerIleTTPLeuTyrSerValProGlnGlnHisIleGlnAspIleThr 226
Db      4281 GGAATCCATCGCAGGTATAGTCTTACACAGCCGAAACACCTGGGTGACATATTCAC 4340
Qy      227 ValValThrAspVal-----AspGlyAspAsnGlyLeuIleAsnTyrGlnValGlnVal 244
Db      4341 GTGGTACCGCATGTGCGGCAAGCTGTAAACCGGCTGTGTGATGCGACGCTG--GTG 4397
Qy      245 AlaAsnGlnThrGlyGlnIleSerValIleAspGlnAspGlyValAlaIleVal 264
Db      4398 GCCAAT-----GGTATGTACGCTTGAACGCTGATGCGGATCAACAGGTGTT 4448
Qy      265 AlaLysAlaSerGlyValAlaGlnGlyThrValThrIleProSerValLysLeuTTPGlnPro 284
Db      4449 GCAATCGACCAAGGACACTAGCGGACCTTGGCAAGTGGTAATCCGCACTCGGACACG 4508
Qy      285 GlyAlaAlaIleTyrLeuTyrGlnLeuGlnValAsnIleValGlySerSerGlyAspValVal 304
Db      4509 GGTGAAGGTATCTTATGAATGTGCGTC-----ACAGCCMAAGCCAGACAGTGT 4562
Qy      305 AspThrTyrAsnLeuAlaThrGlyValArgThrValLysValAlaGlySerGlnPheLeu 324
Db      4563 GATATCTACCGCTTCGCGGTGGGACATCCGCTAGTGGACGTGAAGGCGAAGATCTCTG 4622
Qy      325 IleAsnGlyLysProPheTyrPheThrGlyPheGlyLysHisGlnAspThrAlaValArg 344
Db      4623 ATTAACCAACAAACCGTTCACTTACTGCTTGGTGTGCTATGAAGATCGGACCTTACGT 4682
Qy      345 GlyLysGlnHisAspProAlaTyrMetValHisAspPheGlnLeuMetLysTTPIleGly 364
Db      4683 GGCAAAGATTCGATTAACGTGCTGATGTCAGACACACGCAATTAATGAGCTGATGGG 4742
Qy      365 AlaAsnSerPheArgTyrSerHisTyrProTyrAlaGlnGlnValAlaMetAspPheAlaAsp 384
Db      4743 GCGAATCTCTACCGTACCTCGCATTAACCTTACGCTGAAGATGCTGATCGGACAT 4802
Qy      385 ArgAsnGlyIleValValIleAspGlnThrProAlaValGlyLeuAsnIleAlaLeu-- 403
Db      4803 GAACATGCGCATCGTGGTATGATGAACCTGCTGCGCTTAACTCTCTTTAGGC 4862
Qy      404 MetGlyValSerGlnSerGlyAlaPro--GlnThrPheThrProAspAlaIleAsnAsp 422
Db      4863 ATGTTGTTGGAAGCGGCAACAGCCGAAAGACTGTACACCGAAGACGCTCAACGGG 4922
Qy      423 LysThrGlnGlnAlaHisLysGlnAlaIleArgGlnLeuIleAlaArgAspLysAsnHis 442
Db      4923 GAACCTGCAAGAGCGCACTTACAGCGCATTAAGAGCTGATAGCGCGGTGACAAACAC 4982
Qy      443 AlaSerValValMetTTPSerIleAlaAsnGlnProAlaSerHisGlnAspGlyValArg 462

```

```
Db 4983 CCAAGCGTGGTGTATGGAGTATTGCCAAGAACCGGATACCGTCTCT-----GCACGG 5036
Qy 463 GluTyrPhe-----GluProLeuThrAsnLeuThrArgGlnLeuAspProThiArgPro 480
Db 5037 GAATATTTTCGGCATTTTCGCCATTCGGGAGCAACCGGTAACTCGACCCGACCGGTCCG 5096
Qy 481 IleThrPheAlaAsnValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPhe 500
Db 5097 ATACCTGCGCTCAATGTAATGTTCTGCGACGCTCACACCGATACATCAGCGATCTCTTT 5156
Qy 501 AspValSerCysIleAsnArgTyrPheGlyTyrTyrSerGlnThrGlyAspLeuGlu 520
Db 5157 GATGTCTGTGCTGTAACCGTTATTCAGATGATGTCCAAAGCGCGATTTGGAAACG 5216
Qy 521 AlaGluAlaAlaLeuGluLysGluLeuHisGlyTyrGlnGluLysPheHisArgProIle 540
Db 5217 GCAGAGAGGTACTGGAAAAAGAACTTCTGCGCGAGAGAAACTGCATCAGCCGATT 5276
Qy 541 ValMetThrGluTyrGlyAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuPro 560
Db 5277 ATCATCACCGAATACCGCGTGGATACGTTAGCCGGGCTGCATCAATGTACACCGACATG 5336
Qy 561 TrpSerGluGluPheGlnValGlnMetLeuAspMetTyrHisArgValPheAspArgIle 580
Db 5337 TGGAGTGAAGAGTATCAGTGTGATGCTGATGATGATCATCCGCGCTTTGATCGGCTC 5396
Qy 581 GluSerMetAlaGlyGluHisValTyrAsnPheAlaAspPheGlnThrAsnLeuGlyIle 600
Db 5397 AGCGCGCTGCTCGGTAAACAGTATGGAATTTGCCGATTTTGCACCTCGCAAGGCATA 5456
Qy 601 IleArgValAspGlyAsnLysLysGlyValPheThrArgAspArgLysProLysAlaAla 620
Db 5457 TTGCGGCTTGCGCGTAAAGGATCTTCACTGCGACCGCAACCGAAGTCGGCG 5516
Qy 621 AlaHisSerLeuArgAlaArgTyrThrSerIleAsp 632
Db 5517 GCTTTTCTGCTGCATAAACGCTGACCTGCATGAAC 5552
```

Search completed: March 22, 2005, 20:44:17
Job time : 407 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: March 22, 2005, 14:08:15 ; Search time 781 Seconds
(without alignments)
4805.526 Million cell updates/sec

Title: US-10-757-093-4
Perfect score: 3354
Sequence: 1 MKFLTGLSLTSLAPSLGTP.....RKPKAAASHLRARWTSIDKN 624

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues
Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+g2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10757093/runat_18032005_164455_27758/app_query.fasta_1.775
-DB=N.Geneseq.16dec04 -QFMT=faetap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=bloms62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTMT=ptp -NCRW=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10757093.@CGN_1.1.708.@runat_18032005_164455_27758 -NCPU=6 -ICPU=3
-NO_MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-DEV TIMEOUT=120 -WANT TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: N.Geneseq.16dec04:.*
2: geneseqn1980s:.*
3: geneseqn1990s:.*
4: geneseqn2000s:.*
5: geneseqn2001as:.*
6: geneseqn2002as:.*
7: geneseqn2002bs:.*
8: geneseqn2003as:.*
9: geneseqn2003bs:.*
10: geneseqn2003cs:.*
11: geneseqn2003ds:.*
12: geneseqn2004as:.*
13: geneseqn2004bs:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1756	52.4	1812	3	AAA07939
2	1717.5	51.2	3451	8	ABT16609
3	1717.5	51.2	3451	10	ACC44711
4	1717.5	51.2	14627	8	ABT16610
5	1717.5	51.2	14627	10	ACC44712

6	1717	51.2	4299	3	AAA27980	AAa27980 Vector MS
7	1714.5	51.1	3169	2	AAI38397	AaI38397 pTc 99 p
8	1714.5	51.1	8654	4	AAFP0288	AaFP0288 Nucleotid
9	1714.5	51.1	8654	4	AAFP0285	AaFP0285 Nucleotid
10	1714.5	51.1	8987	4	AAFP0296	AaFP0296 Nucleotid
11	1714.5	51.1	9390	4	AAFP0292	AaFP0292 Nucleotid
12	1714.5	51.1	9390	4	AAFP0294	AaFP0294 Nucleotid
13	1714	51.1	5919	4	AAFP0980	AaFP0980 pSP-GUS
14	1714	51.1	7510	12	ADMO1287	AdMO1287 Plasmid p
15	1714	51.1	8327	12	ADMO1291	AdMO1291 Plasmid p
16	1714	51.1	10011	12	ADMO1737	AdMO1737 Plasmid t
17	1714	51.1	12225	6	ADBO6400	AdBO6400 pLW-48 pl
18	1714	51.1	12225	10	ADBO6532	AdBO6532 Plasmid t
19	1714	51.1	32798	6	ABR97684	ABa97684 Replicati
20	1714	51.1	32798	13	ADP79484	ADp79484 Adenovitr
21	1713.5	51.1	6795	3	AAFC6821	AaFC6821 Reporter
22	1713.5	51.1	6795	3	AAFC6822	AaFC6822 Reporter
23	1713.5	51.1	6795	3	AAFC6817	AaFC6817 Reporter
24	1712.5	51.1	3035	2	AAV37748	AaV37748 PAT1 Gene
25	1712.5	51.1	1809	6	ABR87080	ABr87080 uida DNA
26	1711	51.0	1812	6	ABLe1371	ABl61371 GUS DNA #
27	1711	51.0	1812	8	ABT16606	ABt16606 Artificial
28	1711	51.0	1812	10	ACC44709	Acc44709 E. coli b
29	1711	51.0	1812	13	ADG46099	AdG46099 Bacterial
30	1711	51.0	3824	2	AAV37749	AaV37749 PAT1 gene
31	1709	51.0	1812	4	AAH22503	AaH22503 PCR gene
32	1709	51.0	1812	6	AAH29543	AaH29543 uida DNA
33	1709	51.0	4652	10	ADD27985	Add27985 Beta-gluc
34	1709	51.0	8617	3	AAFC6819	AaFC6819 Reporter
35	1709	51.0	8617	3	AAFC6818	AaFC6818 Reporter
36	1706	51.0	8684	3	AAFC6820	AaFC6820 Reporter
37	1706	50.9	5390	10	ADD27990	Add27990 Caleosin/
38	1706	50.9	5418	10	ADD27987	Add27987 Oleosin/b
39	1703	50.8	11978	6	ABBS3598	ABbs3598 Neureoth
40	1703	50.8	12438	6	ABBS3599	ABbs3599 Plasmid p
41	1700	50.7	1814	10	ADCO1866	ADc01866 Enterohae
42	1697	50.6	1812	6	ABLe1358	ABl61358 GUS DNA #
43	1692	50.4	8012	2	AAK57305	AaK57305 Sugar bee
44	1692	50.4	8012	2	AAK57309	AaK57309 Sugar bee
45	1692	50.4	8798	2	AAK57308	AaK57308 Sugar bee

ALIGNMENTS

RESULT 1	AAA07939	standard; DNA; 1812 BP.
ID	AAA07939	
XX	AAA07939;	
AC	AAA07939;	
DT	26-JAN-2001	(first entry)
XX		
DE	Salmonella beta-glucuronidase gene.	
XX		
KW	Microbial; beta-glucuronidase; GUS; Enterobacter; Salmonella;	
KW	Pseudomonas; Staphylococcus; Thermotoga; transgenic plant; bioindicator;	
KW	transgenic insect; marker; glucuronide detoxification; ds.	
XX		
OS	Salmonella sp.	
XX		
PN	WO20005533-A1.	
XX		
PD	21-SEP-2000.	
XX		
PF	16-MAR-2000; 2000WO-US007107.	
XX		
PR	17-MAR-1999; 99US-00270957.	
XX		
PA	(CAMP-) CAMBIA BIOSYSTEMS LLC.	
XX		
PI	Jefferson RA, Mayer JE;	
XX		
DR	WPI; 2000-647075/52.	

DR P-PSDB; AAB28409.

XX Novel microbial beta-glucuronidase genes and gene products used as
PT reporter/effector molecule, as diagnostic tool, in positive selection, to
PT target molecules to specific cells and to detect and track linked genes.

XX Example 3; Fig 16; 116pp; English.

XX The present sequence encodes a microbial beta-glucuronidase (GUS)
CC protein. GUS genes were obtained from six different genera:
CC Enterobacter/Salmonella, Pseudomonas, Salmonella, Staphylococcus and
CC Thermotoga. Microbial GUS can be used as a reporter/effector molecule for
CC transgenic constructions and in in vitro diagnostic applications. It may
CC also be used to generate sentinel plants that serve as bioindicators of
CC environmental status. It may be used to facilitate the development of a
CC tracking insect populations or to facilitate the development of a
CC bioassay for compounds that affect molecules critical for insect
CC development (e.g. juvenile hormone). Secreted GUS may also serve as a
CC marker for beneficial fungi destined for release into the environment. In
CC animal systems, secreted GUS may be used to achieve extracellular
CC detoxification of glucuronides (e.g. toxin glucuronide) and to examine
CC conjugation patterns of glucuronides. Microbial GUS may also be used in
CC traditional medical diagnostic assays, for drug testing, pharmacokinetic
CC studies, bioavailability studies, diagnosis of diseases and syndromes,
CC following progression of disease or its response to therapy. Microbial
CC GUS has increased thermal stability, high turnover number and enzymatic
CC activity. It is highly specific for the substrate and water soluble, and
CC the substrates are stable

XX Sequence 1812 BP; 437 A; 465 C; 531 G; 379 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1,21e-142	Length:	1812
Score:	1756.00	Matches:	329
Percent Similarity:	71.6%	Conservative:	99
Best Local Similarity:	55.11%	Mismatches:	159
Query Match:	52.36%	Indels:	10
		Gaps:	6

US-10-757-093-4 (1-634) x AAA07939 (1-1812)

```

QY 40 ValArgProGlnArgThrSerSerArgGluLeuValAsnLeuAspGlyLeuTrpLysPhe 59
DB 4 TTAAGTCTGTCTCGAAACCGGAGCGCGAAGATCAAAAAGTGGACGGCTGTGTGTTT 63
QY 60 AlaLeuAspGlyLeuAsnAspThrAlaGlnPro---TrpThrAlaProLeuProLys 78
DB 64 TGTATGATAGCGAAGATCGCGCAACGCCAGCAATGTGCGTCAACCGTTAACCCCA 123
QY 79 GlyLeuGlnCysProValProAlaSerTyrAsnAspIlePheIleSerArgGluIleHis 98
DB 124 AGCCGGCGTATCGCGGTTCCGGAGAGCTATACATGATGTTGCGCTGCCGAGATCCG 183
QY 99 AspHisValGlyTrpValTyrTyrGlnArgGluValIleValProLysGlyTyrPserGln 118
DB 184 AATTATTTGGCAACGTCGTGTATCACCGTGAATAGACATCCGGAAGGCTGGGATCCG 243
QY 119 GluArgTyrLeuValArgAlaGluSerAlaThrHisIleGlyArgIleTyrValAsnAsn 138
DB 244 CAGGCGATAGTGTGCGCTTGTATGATGATGATGATGATGATGATGATGATGATGATG 303
QY 139 ArgLeuValAlaGluHisValGlyTyrThrProGluValAlaPheValThrGluLeu 158
DB 304 CAATTTTAAATGAGAACATCAGGGCGGCTACACGCGTTTGAAGGGATATACAGCACCTT 363
QY 159 ValAlaProGlyGluLysPheArgLeuThrIleGlyValAsnGluLeuThrHisGlu 178
DB 364 ATCTCCCGCGGGAATCCGTCGTATCACCGTATGCGTGAATACGAGCTGAAGTGGCAG 423
QY 179 ThrIleProGlyLysIleThrThrGlyAsnAlaThrGlyLysArgIleGlnThrTyr 198
DB 424 ACGATCCCGCGGCGGCTGTGTGACCCAGGCGC---GTAAACGGTAAAGAACAGCAAGCCTAT 480

```

```

QY 199 GlnHisAspPheTyrAsnTyrAlaGlyLeuAlaArgSerIleTrpLeuTyrSerValPro 218
DB 481 TTCATGATTTCTTTTACACCGCGTATTCATCGCAGGCTATGTCGACCAACCGCG 540
QY 219 GlnGlnHisIleGlnAspIleThrValValThrAspValAspGlyAspAsnGlyLeu 238
DB 541 AAAACTTTTGTGGAAGATATATACCGTCGACGCGAGGTGTCGACGAT-----CTGGCT 594
QY 239 AsnTyrGluValGluValAlaAsnGlnThrThrGlnIleGlnIleSerValIleAsp 258
DB 595 CAGGCTACCGTCCCTCGCGAGTACGCGGAGATGCGCAAGTCCGTAGCTAGCTATGAC 654
QY 259 GluAspGlyAlaIleValAlaLysAlaSerGlyAlaGlnGlyThrValThrIleProSer 278
DB 655 GCGAGAACACACTTGTGCTTCGCGGCAAGGAAAGAGTGAAGTGTGCTGAAGGCG 714
QY 279 ValLysLeuTrpGlnProGlyAlaAlaTyrLeuTyrGlnLeuGlnValAsnIleValGly 298
DB 715 CCGCGCTGTGCGAGCTGCGAGGCGTATCTTTATGAACTCGGGGTC-----ATCGCG 768
QY 299 SerSerGlyAspValValAspThrTyrAsnIleAlaThrGlyValArgThrValLysVal 318
DB 769 CAGCATCAGACGACGAGATGAATATCCGCTGCGCGTGTATTCCTCGGTAGAGTA 828
QY 319 AlaGlySerGlnPheLeuIleAsnGlyLysProPheTyrPheThrGlyPheGlyLysHis 338
DB 829 AAAGGGACACATCTCTGATCAACCATAGCCTTTCTATTATCCGCGGTGCGAGCTCAT 888
QY 339 GluAspThrAlaValArgGlyLysGlyHisAspProAlaTyrMetValHisAspPheGln 358
DB 889 GAAGATGCGCATCTGCGCGTAAAGGTTTGTATTAACGTGCTGATGATGATGATGATGATG 948
QY 359 LeuMetLysTrpIleGlyValAsnSerPheArgThrSerHisIleTyrProTyrAlaGlnIu 378
DB 949 CTATAGACTGTAGTGTGCGTGAACCTTATCCGTACTCGCATTCCTTATCCGTAAGAG 1008
QY 379 ValMetAspPheAlaAspArgAsnGlyIleValIleAspGluThrProAlaValGly 398
DB 1009 ATGCTGACATGGGCGGAGCAACATGCGATCGATCATGTATGATAAGCGCCCGCTCGGA 1068
QY 399 LeuAsnIleAlaLeuMetGlyValSer-----GluSerGlyValAspGlnThrPhe 415
DB 1069 TTCACATCTGCTTTA---GGGATTTAGCTTTGATGTCGCGGAAAAACCAAGAGCTTAC 1125
QY 416 ThrProAspAlaIleAsnAspLysThrGlnGluAlaHisIleGlnAlaIleArgGluLeu 435
DB 1126 AGCGATGAGGCGGTGAACATGAACGACGCGCGCACCCTGCAAGCATTTAAGAGCTG 1185
QY 436 IleAlaArgAspLysAsnHisAlaSerValValMetTrpSerIleAlaAsnGluProAla 455
DB 1186 ATTGCCCGGATAGAACCAACCAACGATCGATGTGAGTATGCGCAACGAAACCGGAT 1245
QY 456 SerHisGluAspGlyAlaArgGlyTyrPheGluProLeuThrAsnLeuThrArgGlnLeu 475
DB 1246 ACCGCGCCGAAACGCGCGCGCAATCTTCGCTCGCCTGCGCAGGCAACGCGCAATCTC 1305
QY 476 AspProThrArgProIleThrPheAlaAsnValIleThrAlaThrTyrGlnLeuAspArg 495
DB 1306 GATCTACACGTCGTAACCTCGCGAACGTAATGTTCTGCGATCGGAAACGCAACAC 1365
QY 496 IleSerAspLeuPheAspValSerCysIleAsnArgTyrPheGlyTyrTrpSerGlnThr 515
DB 1366 ATTACGAGTCTCTTATGTCGTTGTCGTAACCGCTACCGGCTGTATGATGACAAAC 1425
QY 516 GlyAspLeuGlnGluAlaGluAlaIleGluGlyLysGlnLeuHisGlyTyrTrpGlnGluLys 535
DB 1426 GCGGATCTGAGAAAGCTGAGAAAGTGTGAGAAAGACTTCTGCGCTGCGCAAGGAGAA 1485
QY 536 PheHisArgProIleValMetThrGluTyrGlyAlaAspThrLeuAlaGlyLeuHisSer 555
DB 1486 CTCACCGCGCGGCTTATCATCACCGAATACGCGGTGATGACGTTCCAGGCTGCAATTC 1545
QY 556 IleLeuGlyLeuProTrpSerGlnGluPheGlnValGlnMetLeuAspMetTyrHisArg 575

```

Db 1546 ATGTACACGATATGTGACGGAAGTACACAGCGCTGTGATATGATGCATCGC 1605
Qy 576 ValPheAspArgIleGluSerMetAlaGlyValHisValTrpAsnPheAlaAspPheGln 595
Db 1606 GTCTTTGATGTGCTGACGCGCTGCTGCGGACAGGTATGAATCTGCGGCTTCCGCC 1665
Qy 596 ThrAsnLeuGlyIleIleArgValAspGlyAsnLysIleGlyValPheThrArgAspArg 615
Db 1666 ACTTCGACGGGCTTATGCGCGTGGCGGCAAAAAGCTATATTCACCCCGGACAGA 1725
Qy 616 LysProLysAlaAlaAlaHisSerLeuArgAlaArgTrpHisSerIleAsp 632
Db 1726 AAACCAAAATCGCGCGCTTCCTGCTGCAAAAACGCTGACCGCGCATGAC 1776
RESULT 2
ID ABT16609 standard; DNA; 3451 BP.
XX AC ABT16609;
XX DT 03-APR-2003 (first entry)
XX DE Artificial plant chromosome related plasmid DNA SEQ ID No 20.
XX KM Plant artificial chromosome; PAC; transgenic plant; vaccine;
XX KM blood factor; herbicide; stress; agronomical; nutrient quality;
XX KM bacterial artificial chromosome; BAC; Yeast artificial chromosome; YAC;
XX OS ds.
XX Unidentified.
XX PN MO200296923-A1.
XX PD 05-DEC-2002.
XX PF 30-MAY-2002; 2002MO-US017451.
XX PR 30-MAY-2001; 2001US-0294687P.
XX PR 04-JUN-2001; 2001US-0296329P.
XX PA (CHRO-) CHROMOS MOLECULAR SYSTEMS INC.
XX PA (AGRI-) AGRISOMA INC.
XX PI Perez C, Fabijanek SF, Perkins E;
XX MPI; 2003-140436/13.
XX PT Producing artificial chromosome by introducing a nucleic acid into plant
XX PT cell, selecting artificial chromosome that has one or more repeat regions
XX PT with equivalent amounts of euchromatic and heterochromatic nucleic acids.
XX PS Disclosure; Page 248-249; 2699p; English.
XX CC The invention relates to a novel method for producing plant artificial
XX CC chromosomes. The invention also relates to methods for targeting
XX CC insertion of heterologous DNA into plant artificial chromosomes, methods
XX CC for delivery of plant chromosomes to selected cells and tissues. The
XX CC isolated plant artificial chromosome (PAC) is useful for producing a
XX CC transgenic plant, which involves introducing the PAC into a plant cell.
XX CC The PAC comprises a heterologous nucleic acid encoding a gene product
XX CC such as enzymes, antisense RNA, rRNA, rDNA, structural proteins, marker
XX CC proteins, ligands, receptors, ribozymes, therapeutic proteins, and
XX CC biopharmaceutical proteins, vaccines, blood factors, antigens, hormones,
XX CC cytokines, growth factors, antibodies, or a product that provides for
XX CC resistance to diseases, insects, herbicides, or stress in a plant. The
XX CC heterologous nucleic acid optionally encodes a product that provides an
XX CC agronomically important trait in the plant, e.g. a product that alters
XX CC nutrient use and/or improves the nutrient quality of the plant. The
XX CC heterologous nucleic acid is contained within a bacterial artificial
XX CC chromosome (BAC) or a yeast artificial chromosome (YAC). This
XX CC polynucleotide sequence represents the DNA of a plasmid used in the
XX CC method of the invention

XX SQ Sequence 3451 BP; 951 A; 707 C; 851 G; 942 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 6,15e-139 Length: 3451
Score: 1717.50 Matches: 338
Percent Similarity: 68.84% Conservative: 95
Best Local Similarity: 53.74% Mismatches: 175
Query Match: 51.21% Indels: 21
DB: Gaps: 11
US-10-757-093-4 (1-634) x ABT16609 (1-3451)
Qy 13 AlaAlaProSerIleuGlyThrProAlaAlaArgHisPheProArgAsnGlu----- 29
Db 1221 TCTGCACCGGATTCGAGATC-----GAATTCGCCGCGCGCGGAAATTCAGTAC 1268
Qy 30 ---MetThrGlnHisGluGlnProLeuIleLysValArgProGlnArgThrSerSerArg 48
Db 1269 TGGATCCCGCGGATCGATGCTTATG---TTACGTCCTGTAGAAACCCCAACCGGT 1325
Qy 49 GlnLeuValAsnLeuAspGlyLeuTrpLysPheAlaLeu---AlaSerGlyLeuAsnAsp 67
Db 1326 GAATATCAAAAACCTGCACGCGCTGTGGCATTGCTGATGCGCAAAACCTGGAATT 1385
Qy 68 ThrAlaGlnProThrAlaProLeuProLysGlyLeuGluCysProValProAlaSer 87
Db 1386 GACACGCTGTGGTGGAAAGCGGCTTACAAAGAACCGCGCAATTCGTGCGACGACGT 1445
Qy 88 TyrAsnAspIlePheIleSerArgGluIleHisAspHisValGlyTrpValTyrTrpGln 107
Db 1446 TTTAACGATCAGTTCGCCGCGATGACATATTCGAATATGTTGGCAACGCTGATACAG 1505
Qy 108 ArgGluValIleValProLysGlyTrpSerGlnGluArgTrpLeuValArgAlaGluSer 127
Db 1506 CGCGAAGCTTTATACCGAAAGGTTGGGACGCGACGATACGTCGCTTCGATGCG 1565
Qy 128 AlaThrHisHisIleGlyIleTyrValAsnAsnArgLeuValAlaGlnHisValGlyGly 147
Db 1566 GTACATCATTTACCGCAAGAGTGGGTCAATATACAGAAAGTGTATGACATCAGCGCGC 1625
Qy 148 TyrThrProPheGlnAlaAspValThrGluLeuValAlaProGlyGluLysPheArgLeu 167
Db 1626 TATAGCCCATTTGAACCCGATGTCACGCGTATGTTATGCGGAAAGTGAACGATC 1685
Qy 168 ThrIleGlyValAsnAsnGluLeuThrHisGluThrIleProProGlyLysIleThrThr 187
Db 1686 ACAGTTTGTTGTAACACGAACTGAACCTGACAGACTATCCCGCGGAAATGGATTAAC 1745
Qy 188 GlyAsnAlaThrGlyLysArgIleGlnThrTyrGlnHisAspPheTyrAsnTyrAlaGly 207
Db 1746 ---GACGAAAACCGCAAGAAAAGACGCTTACTTCCATGATATCTTAACTACGCCGGG 1802
Qy 208 LeuAlaArgSerIleTrpLeuTyrSerValProGlnGlnHisIleGlnAspIleThrVal 227
Db 1803 ATCCATCGCAGCGTATATGCTTACACACCGCAACACTCGGTGACATATCACCGTG 1862
Qy 228 ValThrAspVal-----AspGlyAspAsnGlyLeuIleAsnTyrGluValAla 245
Db 1863 GTACACCAATGCGCGCAAGCTGTATACACGCGCTGTTGACACGACGAGTG---GTGGCC 1919
Qy 246 AsnGlnThrThrGlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleValAla 265
Db 1920 AAT-----GATATGACGCTTGAACCTCGTAATCGGATCAACAGGTGGTGGCA 1970
Qy 266 LysAlaSerGlyValaGlnGlyThrValThrIleProSerValLysLeuTrpGlnProGly 285
Db 1971 ACTGCAACAGCACCGCGGACTTTCGCAAGTGTATATCCGACCTGTGGCAACGGGGT 2030
Qy 286 AlaAlaTyrIleuTyrGlnLeuGlnValaAsnIleValaGlySerSerGlyAspValValAsp 305
Db 2031 GAAGTTATCTATGTAACGTACGTC-----ACAGCCAAAAGCCAGACAGAGTGTAT 2084

```

QY 306 ThrTyrAsnLeuAlaThrGlyValArgThrValIleValIleGlySerGlnPheLeuIle 325
DB 2085 ATCTACCCGCTGGCGCGCGCATCCGCTCAGTGCAGTGAAGGCCAAGCATTCTGCATC 2144
QY 326 AsnGlyLysProPheTyrPheThrGlyPheGlyLysHisGluAspThrAlaValArgIly 345
DB 2145 AACCAACAAACGTTCTACTTACTTACGCTTGGCCGTCATGAGATGGCATTTGGCGGGC 2204
QY 346 LysGlyHisAspProAlaTyrMetValHisAspPheGlnMetLysTrpIleGlyAla 365
DB 2205 AAAGGATTCATTAACGTCGATGTCGACGATCAACGATTAATGACCTGGAATGGGGCC 2264
QY 366 AsnSerPheArgThrSerHisTyrProTyrAlaGluGluValMetAspPheAlaAspArg 385
DB 2265 AACTCTCACTGCTACCTGCACTTACCTTACGCTGAGAGATGCTGCACTGGCAGATGAA 2324
QY 386 AsnGlyIleValIleAspGluThrProAlaValIleLysAsnIleAlaLeu--Met 404
DB 2325 CATGGCATCGTGGATGATGATGAAGATGACAGCTGTCGCTTAACTCTCTTTAGGCAT 2384
QY 405 GlyValSerGluSerGlyAlaPro--GlnThrPheThrProAspAlaIleAsnAspLys 423
DB 2385 GATTTCGAAACGGCGCAACAAAGGAAAGGAACTGTACACGAGAGAGCATCAACGGGGAA 2444
QY 424 ThrGlnGluAlaHisLysGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHisAla 443
DB 2445 ACTCAGACGGCGCACTTACAGCGCATTAAGAGCTGATGCGGTGACAAACCCACCA 2504
QY 444 SerValValMetTyrSerIleAlaAsnGluProAlaSerHisGluAspGlyAlaArgIly 463
DB 2505 AGCGGTGTATGTGAGTATGTCACCAACGAAACCGGATCCGTCGCAAGGTGACAGGAA 2564
QY 464 TyrPheGluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPhe 483
DB 2565 TATTTGCGCGCACTGGCGGAGAACGCGTAACTGCATCCGACGCTCGCATCACCCTGC 2624
QY 484 AlaAsnValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPheAspValSer 503
DB 2625 GTCAATGTATATGTTTCGCGACGCTCACACCGATACCATCGCATCTCTTATGATGCTG 2684
QY 504 CysIleAsnArgTyrPheGlyTyrPyrSerGlnThrGlyAspLeuGluAlaIleAla 523
DB 2685 TGCTCTGAACCGTTATATACGTTGATGTCACAAACCGCGCATTTGGAAACGCGAGAA 2744
QY 524 AlaLeuGlnLysGluLeuHisGlyTyrGlnGluLysPheHisArgProIleValMetThr 543
DB 2745 GTACTGAAAAAGAACTTCTGGCTGCGACGAGAAACTGCATACGCGATATATCACC 2804
QY 544 GlyTyrGlyAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTyrPheSerGlu 563
DB 2805 GAATACGGCGCTGATACGTTACCGCGCTGCACTCAATGATACACCGACATGTGAGTGA 2864
QY 564 GluPheGlnValGlnMetLeuAspMetTyrHisArgValPheAspArgIleGluSerMet 583
DB 2865 GAGTATACAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2924
QY 584 AlaGlyGlnHisValTyrAsnPheAlaAspPheGlnThrAsnLeuGlyIleIleArgVal 603
DB 2925 GTGCGTGAACAGGATGAGAAATTTGCGCATTTGGCGACCTTCGCAAGGCATATTTGGCG 2984
QY 604 AspGlyAsnLysLysGlyValPheThrArgAspArgLysProLysAlaIleHisSer 623
DB 2985 GGGCGTTACAAAGAGGGGATCTTCAACCGCGACCGCAACCGAAGTGGCGGCTTTCTG 3044
QY 624 LeuArgAlaArgTyrThrSerIleAsp 632
DB 3045 CTGCAAAAAACGTGACTGCGCATGAAAC 3071

```

```

XX 29-MAY-2003 (first entry)
XX DT
XX XX Plasmid pAg1a nucleotide sequence SEQ ID NO:108.
DE XX
XX XX Chromosome-based platform; artificial chromosome; eukaryotic chromosome;
XX KM at site; integrase; recombinase; Acs; gene therapy; transgenic animal;
XX KM platform artificial chromosome expression system; gene; ds.
XX OS
XX XX Synthetic.
XX PN
XX MO200297059-A2.
XX PD
XX 05-DEC-2002.
XX PF
XX 30-MAY-2002; 2002MO-US017452.
XX PR
XX 30-MAY-2001; 2001US-0294758P.
XX PR
XX 21-MAR-2002; 2002US-0366891P.
XX XX
XX PA (CHRO-) CHROMOS MOLECULAR SYSTEMS INC.
XX PI Perkins E, Perez C, Lindenbaum M, Greene A, Leung J, Fleming E;
XX PI Stewart S, Shellard J;
XX DR MPI; 2003-140461/13.
XX PT
XX PT Novel eukaryotic chromosome comprising one or many att sites which
XX PT permits site-directed integration in the presence of lambda-integrase,
XX PT useful for site-specific recombination-directed integration of DNA of
XX PT interest.
XX PS
XX XX Example 12; Page 234-235; 272pp; English.
XX CC
XX CC The present invention describes a eukaryotic chromosome (1) comprising
XX CC one or several att sites, where an att site is heterologous to the
XX CC chromosome, and permits site-directed integration in the presence of
XX CC lambda-integrase. Also described: (1) a platform artificial chromosome
XX CC expression system (Aces) (11) comprising several sites that participate
XX CC in recombinase catalyzed recombination; and (2) a method (M1) for
XX CC introducing a heterologous nucleic acid into a platform artificial
XX CC chromosome. (1) can be used in gene therapy. (M1) is useful for
XX CC introducing a heterologous nucleic acid molecule into a platform
XX CC artificial chromosome, preferably an Aces. (11) is useful for producing a
XX CC transgenic animal (e.g. a fish, insect, reptile, amphibian, arachnid, or
XX CC mammal) by introducing (11) by cell fusion, lipid-mediated transfection,
XX CC by a carrier system, microinjection, microcell fusion, electroporation,
XX CC microprojectile bombardment or direct DNA transfer into an embryonic
XX CC cell, preferably a stem cell or an embryo. (11) comprises a heterologous
XX CC nucleic acid that encodes a therapeutic product which is useful for
XX CC making a library of Aces comprising random portions of a genome. ACC44612
XX CC to ACC44732 and ABP96650 to ABP96657 represent sequences used in the
XX CC exemplification of the present invention
XX CC
XX SQ Sequence 3451 BP; 951 A; 707 C; 851 G; 942 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: 6,15e-139 Length: 3451
XX Score: 1717.50 Matches: 338
XX Percent Similarity: 68.84% Conservative: 95
XX Best Local Similarity: 53.74% Mismatches: 175
XX Query Match: 51.21% Indels: 21
XX DB: Gaps: 11

```

```

US-10-757-093-4 (1-634) x ACC44711 (1-3451)
QY 13 AAlaAProSerLeuGlyThrProAlaAlaArgHisPheProArgAsnGlu----- 29
DB 1221 TCTGCAACCGGATCTCGAATC-----GAATTCCTCCGCGCGCGGGAATTCCTAG 1268
QY 30 --MetThrGlnHisGlnGlnProLeuIleLysValArgProGlnAlaArgThrSerSerArg 48
DB 1269 TGGATCCCGCGGATCGGTCAAGTCCCTTATG---TTACGTCTGTGAGAAACCCCAACCCGT 1325

```

QY 49 GluLeuValAsnLeuAspGlyLeuTrpLysPheAlaLeu---AlaSerGlyLeuAsnAsp 67
 DB 1326 GAATATCAAAACATCGAGCGCTGGTGGCATTCAGTGGATCGCAAACTGGCAATT 1385
 QY 68 ThrAlaGlnProTrpThrAlaProLeuProLysGlyLeuGluCysProValProAlaSer 87
 DB 1386 GAGCAGCGTTGGTGGAAAGCGGGTTACAGAAAGCGGGCAATTGGTGGCCAGGAGT 1445
 QY 88 TyrAsnAspIlePheIleSerArgGluIleHisAspHisValGlyTrpValIleTyrGln 107
 DB 1446 TTTCAGATCATGTCGGCGATGCGATATTCGTAATATGTCGGCAACGTCGTATCAG 1505
 QY 108 ArgGluValIleValProLysGlyTrpSerGlnGluArgTyrLeuValArgAlaGluSer 127
 DB 1506 CGCAAGCTCTTATACCGAAAGTTGGGCGAGCGGTATCGTGGCTTCGATGCG 1565
 QY 128 AlaThrHisHisGlyArgIleTyrValAsnAsnArgLeuValAlaGluHisValGly 147
 DB 1566 GTCACTCATACGCAAAAGTGGGTCAATATACAGAAAGTGAATGAGCATACAGGGCGC 1625
 QY 148 TyrThrProPheGluAlaAspValThrGluLeuValAlaProGlyGluLysPheArgLeu 167
 DB 1626 TATACCCCATTTGAAGCGATGTCACGCGTATGTTATTCGCGGAAAGTGAACGTATC 1685
 QY 168 ThrIleGlyValAsnAsnGluLeuThrHisGluThrIleProProGlyLysIleThrThr 187
 DB 1686 ACGATTGTGTGAACCAAGCACTGAACCTGACAGCATATCCCGCGGAAAGTGAATACC 1745
 QY 188 GlyAsnAlaThrGlyLysArgIleGlnThrTyrGlnHisAspPheTyrAsnTyrAlaGly 207
 DB 1746 ---GACGAAACCGCAAGAAAGCAAGTCTTACTTCATGATTTCTTAACTACGCCGG 1802
 QY 208 LeuAlaArgSerIleTrpLeuTyrSerValProGlnGlnHisIleGlnAspIleThrVal 227
 DB 1803 ATCCATCGACGCTTAATGCTCTACACCAACCGCAACCTGGGTGAGCATATCACCGTG 1862
 QY 228 ValThrAspVal-----AspGlyAspAsnGlyLeuIleAsnTyrGluValGluValAla 245
 DB 1863 GTACCGCATGTGCGCAAGACTGTAAACCGCTGTGTATCTGGCGAGTG---GTGGCC 1919
 QY 246 AsnGlnThrThrGlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleValAla 265
 DB 1920 AAT-----GCTGATGTACAGCTTGAACCTGCGTATGCGATCAACAGGTGGTCA 1970
 QY 266 LysAlaSerGlyAlaGlnGlyThrValThrIleProSerValLysLeuTrpGlnProGly 285
 DB 1971 ACTGACAAAGCACACAGCGGACTTGTGCAAGTGTGAATCCGCACTCTGGCAACCGGCT 2030
 QY 286 AlaAlaIleTyrLeuTyrGlnLeuGlnValAsnIleValGlySerSerGlyAspValAlaAsp 305
 DB 2031 GAAGGTTATCTTATGAACCTGTACGTC-----ACAGCCAAAGCCAGACAGTGTGAT 2084
 QY 306 ThrTyrAsnLeuAlaThrGlyValArgThrValLysValAlaGlySerGlnPheLeuIle 325
 DB 2085 ATCTACCCGCTGGCGCTGGCATCCGTCAGTGGCACTGAAGGCGCAACAGTTCCTGATC 2144
 QY 326 AsnGlyLysProPheTyrPheThrGlyPheGlyLysHisGluAspThrAlaValArgGly 345
 DB 2145 AACCAACAAACGTTCTTACTTACGCTTGGCGGTATGAGATGCGGATTGGCGGCGC 2204
 QY 346 LysGlyHisAspProAlaTyrMetValHisAspPheGlnLeuMetLysTrpIleGlyVala 365
 DB 2205 AAAGGATTCATTAACGCTGATGTGTGACAGATCACCATTAATATGAGCTGATTTGGGGCC 2264
 QY 366 AsnSerPheArgThrSerHisTyrProTyrAlaGluGluValMetAspPheAlaAspArg 385
 DB 2265 AACTCTCAACCGTACTGCACTTACCTTAACGCTGAAGAGATGCTCGACTGGCGAGATGAA 2324
 QY 386 AsnGlyIleValValIleAspGluThrProAlaValGlyLeuAsnIleAlaLeu--Met 404
 DB 2325 CATGGCATCGTGGTGAATTGATGAAGTGAACGAGCTGTGGCTTTAACTCTCTTAGGCATT 2384

QY 405 GlyValSerGlySerGlyAlaPro---GlnThrPheThrProAspAlaIleAsnAspLys 423
 DB 2385 GCTTTGAACCGCGGCAACAGCCGAAGAACTGTACAGCAAGAGCATGCAAGCGGAA 2444
 QY 424 ThrGlnGluAlaHisLysGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHisAla 443
 DB 2445 ACTCACAGCGCGCATTAACGCGGATTAAGAGCTGATAGCGGTACAAACCAACCCCA 2504
 QY 444 SerValValMetTrpSerIleAlaAsnGluProAlaSerHisGluAspGlyAlaArgGlu 463
 DB 2505 AGCGTGTATGTGGAGTATTTGGCAACGAAACCGGATATCCGTCGCGAGGTGACGGGAA 2564
 QY 464 TyrPheGluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPhe 483
 DB 2565 TATTTGCGCGCATGCGCGGAGCAACGCGTAACTGATCCGACGCGTCCGATCAGCTGC 2624
 QY 484 AlaAsnValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPheAspValSer 503
 DB 2625 GTCAATGTATGTTCGCGACGCTCACACCGATACCATCAGCATCTCTTGTATGTGCTG 2684
 QY 504 CysIleAsnAspArgTyrPheGlyTyrTrpSerGlnThrGlyAspLeuGluAlaGluAla 523
 DB 2685 TGCTGACCGGTATTAACGTTGGTGTATGTCCAAGCGCGGATTTGAACGCGAGAGAG 2744
 QY 524 AlaLeuGluLysGluLeuHisGlyTrpGlnGluLysPheHisArgProIleValMetThr 543
 DB 2745 GTACTGAAAGAAACATTCGCGCTGCGAGAGAACTGCAATGACCGATTAATCATCAC 2804
 QY 544 GluTyrGlyAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTrpSerGlu 563
 DB 2805 GAATACGCGGTGATACGTTAGCGCGGCTGCACTCAATGTACACCGACATGTGAGTGA 2864
 QY 564 GluPheGlnValGlnMetLeuAspMetTyrHisArgValPheAspArgIleGluSerMet 583
 DB 2865 GAGTATCAGTGCATAGCTGGATGTATGTATCACCGGCTCTTATGCTGCACCGCTC 2924
 QY 584 AlaGlyGlnHisValTrpAsnPheAlaAspPheGlnThrAsnLeuGlyIleIleArgVal 603
 DB 2925 GTGCGTGAACAGTATAGAAATTTCCGCGATTTTGCAGCTCGCAAGGCAATATGCGGCT 2984
 QY 604 AspGlyAsnLysLysGlyValPheThrArgAspArgLysProLysAlaAlaHisSer 623
 DB 2985 GCGCGTAAACAAAGAGGAGATCTTACCGCGCAACCGCAAGCGAAGTGGCGCTTTCTG 3044
 QY 624 LeuArgAlaArgTrpThrSerIleAsp 632
 DB 3045 CTGCAAAAGCGTGAACCTGGCATGAAC 3071
 DB
 RESULT 4
 ABT16610
 ID ABT16610 standard; DNA; 14627 BP.
 XX
 AC ABT16610;
 XX
 DT 03-APR-2003 (first entry)
 XX
 DE Artificial plant chromosome related plasmid DNA SEQ ID NO 21.
 XX
 KW Plant artificial chromosome; PAC; transgenic plant; vaccine;
 KW blood factor; herbicide; stress; agronomical; nutrient quality;
 KW bacterial artificial chromosome; BAC; yeast artificial chromosome; YAC;
 KW de.
 XX
 OS Unidentified.
 XX
 PN MO200296923-A1.
 XX
 PD 05-DEC-2002.
 XX
 PF 30-MAY-2002; 2002MO-US017451.
 XX
 PR 30-MAY-2001; 2001US-0294687P.
 PR 04-JUN-2001; 2001US-0296329P.

XX (CHRO-) CHROMOS MOLECULAR SYSTEMS INC.
PA (AGRI-) AGRISOMA INC.

PI Perez C, Fabijanski SF, Perkins E;

XX WPI; 2003-140436/13.

XX Producing artificial chromosome by introducing a nucleic acid into plant
PT cell, selecting artificial chromosome that has one or more repeat regions
PT with equivalent amounts of euchromatic and heterochromatic nucleic acids.

XX Example 5; Page 249-253; 269pp; English.

XX The invention relates to a novel method for producing plant artificial
CC chromosomes. The invention also relates to methods for targeting
CC insertion of heterologous DNA into plant artificial chromosomes, methods
CC for delivery of plant chromosomes to selected cells and tissues. The
CC isolated plant artificial chromosome (PAC) is useful for producing a
CC transgenic plant, which involves introducing the PAC into a plant cell.
CC The PAC comprises a heterologous nucleic acid encoding a gene product
CC such as enzymes, antisense RNA, tRNA, rDNA, structural proteins, marker
CC proteins, ligands, receptors, ribozymes, therapeutic proteins, and
CC biopharmaceutical proteins, vaccines, blood factors, antigens, hormones,
CC cytochromes, growth factors, antibodies, or a product that provides for
CC resistance to diseases, insects, herbicides, or stress in a plant. The
CC heterologous nucleic acid optionally encodes a product that provides an
CC agronomically important trait in the plant, e.g. a product that alters
CC nutrient use and/or improves the nutrient quality of the plant. The
CC heterologous nucleic acid is contained within a bacterial artificial
CC chromosome (BAC) or a yeast artificial chromosome (YAC). This
CC polynucleotide sequence represents the DNA of a plasmid used in the
CC method of the invention

XX Sequence 14627 BP; 3582 A; 3759 C; 3900 G; 3386 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 3,99e-138 Length: 14627
Score: 1717.50 Matches: 338
Percent Similarity: 68.84% Conservative: 95
Best Local Similarity: 53.74% Mismatches: 175
Query Match: 51.21% Indels: 21
DB: Gaps: 11

US-10-757-093-4 (1-634) x ABT16610 (1-14627)

QY 13 AAlaAlaProSerLeuGlyThrProAlaAlaArgHisPheProArgAsnGlu----- 29
DB 12069 TCTGCACCGGATCTCGAGATC-----GAATTCCCGCGCGCGCAATTCTACTAG 12116
QY 30 --MetThrGlnHisGlnProLeuLeuValArgProGlnArgThrSerSerArg 48
DB 12117 TGGATCCCGCGGATCGGTGCTCCCTTATG---TTAGTCTCTGTAAGAACCCCAACCGT 12173
QY 49 GlnLeuValAsnLeuAspGlyLeuTrpLysPheAlaLeu---AlaSerGlyLeuAsnAsp 67
DB 12174 GAATCATAAACCTGCACGCGCTGTGGCATTCAGTCTGGATGCGCAAACTGTGGAAT 12233
QY 68 ThrAlaGlnProTrpThrAlaProLeuProLysGlyLeuGlnLysProValProAlaSer 87
DB 12234 GAGCAGCGTGTGGGAAACCGGCTTACAAAGACCGGCAATTCTGTGCGAGGAGT 12293
QY 88 TyrAsnAspIlePheIleSerArgGluLeuHisAspHisValGlyTrpValTyrTrpGln 107
DB 12294 TTTAACGATACAGTTCGCCGATGCAGATATTCGTAATTATGTGGCAACGCTGTGATCAG 12353
QY 108 ArgGluValIleValProLysGlyTrpSerGlnGluArgTrpLeuValArgAlaGlnSer 127
DB 12354 CGGAGATCTTTATACGAAAGGTGGCGAGCGAGGTATCGTGTGCTTCGATGCG 12413
QY 128 AlaThrHisIleGlyArgIleTyrValAsnAspArgLeuValAlaGlnHisValGlyGly 147
DB 12414 GTCACTCATTAACGCAAGGTGGGTCAATTAATCAGAAAGTGTATGAGATCAGGGCGGC 12473

QY 148 TyrThrProPheGlnAlaAspValThrGluLeuValAlaProGlyGluLysPheArgLeu 167
DB 12474 TATACCCATTGTGAAACCGATGTCACGCCGTATGTTATTCGCCGAAAGTCACTATC 12533
QY 168 ThrIleGlyValAsnAsnGluLeuThrHisGluThrIleProProGlyLysIleThrThr 187
DB 12534 ACAGTTGTGTGTAACAACGAACTGAACCTGGACAGTATCCCGCGGAAATGGTATTAAC 12593
QY 188 GlyAsnAlaThrGlyLysArgIleGlnThrTyrGlnHisAspPheTyrAsnTrpAlaGly 207
DB 12594 --GACGAAACCGGCAAGAAAGAGCTTCACTTCATGATTTCTTAACTACACCGCGG 12650
QY 208 LeuAlaArgSerIleTrpLeuTyrSerValProGlnGlnHisIleGlnAspIleThrVal 227
DB 12651 ATCCATCGACGCTATAGCTTACACACCGGAAACCTGGGTGACATATACCGTG 12710
QY 228 ValThrAspVal-----AspGlyAsnGlyLeuIleAsnTrpGluValGluValAla 245
DB 12711 GTGACCATGTGCGCGAAGCTGTACACAGCGGTGTGTGATCGCAGGTG--GTGGCC 12767
QY 246 AsnGlnThrThrGlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleValAla 265
DB 12768 AAT-----GCTATGTCAAGCGTGAACCTGATGCGGATCAACAGGTGTGGA 12818
QY 266 LysAlaSerGlyValaGlnGlyThrValThrIleProSerValLysLeuTrpGlnProGly 285
DB 12819 ACTGGAACAGCACCGCGGACTTTGCAAGTGTGATACCCGACTCTGGCAACCGGCT 12878
QY 286 AlaAlaTyrLeuTyrGlnLeuGlnValaIleValGlySerSerGlyAspValValaAsp 305
DB 12879 GAAGGTATCTCATGAACTGTACGTC-----ACAGCCAAAGCCAGACAGTGTGAT 12932
QY 306 ThrTyrAsnLeuAlaThrGlyValaArgThrValLysValAlaGlySerGlnPheLeuIle 325
DB 12933 ATCTACCGCTGCGCGTCCGATCGGTCAAGTGAAGGCGCAAGGATTCCTGATC 12992
QY 326 AsnGlyLysProPheTyrPheThrGlyPheGlyLysHisGluAspThrAlaValArgGly 345
DB 12993 AACCAAAACCGTTCATCTTACTGCTTTGGCCGTCATGAAGATCGGATTTGCCCGGC 13052
QY 346 LysGlyHisAspProAlaTyrMetValHisAspPheGlnLeuMetLysTrpIleGlyAla 365
DB 13053 AAAGGATGTGATTAACCTGCTGATGTCAGCATCAACGCTTAATGATGATGATTTGGGCGC 13112
QY 366 AsnSerPheArgThrSerHisTyrProGlyAlaGlnGluValMetAspPheAlaAspArg 385
DB 13113 AACTCTACCGTACCTCGCATTAACCTTACGCTGAAGAGATGCTGACTGGGAGATGAA 13172
QY 386 AsnGlyIleValIleAspGluThrProAlaValaGlyLeuAsnIleAlaLeu--Met 404
DB 13173 CATGGATGTGTGATGATTAAGAACTGCAGCTGTGCTTTAACTCTCTTAAAGCAT 13232
QY 405 GlyValSerGlnSerGlyAlaPro--GlnThrPheThrProAspAlaIleAsnAspLys 423
DB 13233 GGTTCGAAACCGGCAACGAAAGCTGACGAAAGCTGACGGAAGGCACTGCAACGGGGA 13292
QY 424 ThrGlnGluAlaHisLeuGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHisAla 443
DB 13293 ACTCAGACGCGACTTACAGCGATTAAGAGCTGATAGCCGCTGACAAACCAACCCCA 13352
QY 444 SerValaMetTrpSerIleAlaAsnGluProAlaSerHisGluAspGlyAlaArgGln 463
DB 13353 AGCGTGTATGTGAGATTTGCCAAGAACCGGATACCCGTCGGAAGTGTGACGGGAA 13412
QY 464 TyrPheGlnProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPhe 483
DB 13413 TATTTGCGCCCACTGCGGAAACCAACGCTAACTGATCCACGCGTCCGATCACTCTGC 13472
QY 484 AlaAsnValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPheAspValSer 503
DB 13473 GTCAATGTATATGTTTCGACGCTCAACACGATACCATGAGATCTTTGATGTGCTG 13532

Oy	504	Cys11leasnArgTYrPheg1yTTPYrYSerG1nthG1yAspReuG1u1a1a1a	523
Db	13533	TGcCTGAACCGCTTATTACGGTTGGATGTCMAAGCGGGCATTTGGAAACGGCAGAGAA	13592
Oy	524	AlAlenUG1u1ySG1ueuNH1SG1yTTPG1ngu1yAsPheH1sArgPro11eValMetThr	543
Db	13593	GTACTGGAAAAAGAACTTCCTGGCCCTGGCAGAGAAACTGCATCATGCGCGATTTCATCAC	13652
Oy	544	GlU1yTCT1yAlaAsPRT1leuAlaG1y1eNH1SeSer11leuG1y1euePro1TrpSerG1u	563
Db	13653	GAATACGCGCGTGAATACGTTAGCCGGCGCTGCATCATGAACCGACATGTGGAGTGA	13712
Oy	564	GlU1PheG1InValG1InMet1eueAsPMeTYrH1sArgValPheAsPArg11eG1ueSerMet	583
Db	13713	GAGATCATGATGTCATGCTGCATATATATACACCGGCTCTTTGATGCGCTCAGCGCGCT	13772
Oy	584	AlaG1yG1uH1sVal1TTPAsnPh1eAlaAsPPhG1InThAsn1eUG1y11e1eArgVal	603
Db	13773	GTCGGTCAACAGGATGATGGAATTCGCCGATTTTGCACCTCGCAAGGCAATATTGCGCGCT	13832
Oy	604	AsP1yAsn1yS1y1SG1yVal1PheThArAsPArg1yAsPPro1ySAla1a1aH1sSer	623
Db	13833	GCGGATTACAGAAAGGGGATCTTCAACCCGCGACCGCAACCGAAGTCGCGGCTTTCTG	13892
Oy	624	LeuArgAlaArgTTPThSer1eAsp	632
Db	13893	CTGCAGAAACGCTGCATCGCATGAAC	13919
RESULT 5			
ACC44712	ID ACC44712 standard; DNA; 14627 BP.		
XX	ACC44712;		
DT	29-MAY-2003	(first entry)	
XX	Plasmid pAg11a nucleotide sequence SEQ ID NO:109.		
XX	Chromosome-based platform; artificial chromosome; eukaryotic chromosome;		
KW	att site; integrase; recombinase; Acs; gene therapy; transgenic animal;		
KW	platform artificial chromosome expression system; gene; de.		
XX	Synthetic.		
OS	WO200297059-A2.		
PN	05-DEC-2002.		
PD	30-MAY-2002; 2002MO-US017452.		
PE	30-MAY-2001; 2001US-0294758P.		
PR	21-MAR-2002; 2002US-0366891P.		
XX	(CHRO-) CHROMOS MOLECULAR SYSTEMS INC.		
PA	Perkins E, Perez C, Lindenbaum M, Greene A, Leung J, Fleming E;		
XX	Stewart S, Shellard J;		
DR	WPI; 2003-140461/13.		
XX	Novel eukaryotic chromosome comprising one or many att sites which		
PT	permute site-directed integration in the presence of lambda-integrase,		
PT	useful for site-specific recombination-directed integration of DNA of		
PT	interest.		
XX	Example 12; Page 235-239; 272pp; English.		
PS	The present invention describes a eukaryotic chromosome (I) comprising		
CC	one or several att sites, where an att site is heterologous to the		
CC	chromosome, and permits site-directed integration in the presence of		
CC	lambda-integrase, and permits site-directed integration in the presence of		
CC	expression system (Acses) (II) comprising several sites that participate		
CC	in recombinase catalysed recombination; and (2) a method (M1) for		

DB	Seq	Alignment Scores:	Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	DB:	
CC	introducing a heterologous nucleic acid into a platform artificial chromosome. (I) can be used in gene therapy. (M) is useful for introducing a heterologous nucleic acid molecule into a platform artificial chromosome, preferably an Acee. (II) is useful for producing a transgenic animal (e.g. a fish, insect, reptile, amphibian, arachnid, or mammal) by introducing (II) by cell fusion, lipid-mediated transfection by a carrier system, microinjection, microcell fusion, electroporation, microprojectile bombardment or direct DNA transfer into an embryonic cell, preferably a stem cell or an embryo. (II) comprises a heterologous nucleic acid that encodes a therapeutic product which is useful for making a library of Acs comprising random portions of a genome. ACC44612 to ACC44732 and ABP96650 to ABP96657 represent sequences used in the exemplification of the present invention	3.99e-138	Length: 14627	1717.50	Matches: 338	Conservative: 95	Mismatches: 175	Indels: 21	Gaps: 11
US-10-757-093-4	(1-634) x ACC44712 (1-14627)								
OY	13 AAlaAProSeReuEulYThrProAlaAlaArgHisPheProArgAynGlu-----	29							
DB	12069 TCTGCACCGGATCTCGAGATC-----GAATTCGGCGGCGCGGATTCAGTAC	12116							
OY	30 ---MethGlnHisGluGlnProLeuLelyValAlaArgProGlnArgThiSeSerArg	48							
DB	12117 TGAATCCCGGGTACGGTACGTCCTTATG---TTACGCTCGTGAAACCCCAACCGGT	12173							
OY	49 GluLeuValAlaSeruSrgLyeuTrpLysPheAlaLeu---AlaSerGlyLeuAsnAsp	67							
DB	12174 GAAATCAAAAACCTCGACGGCCCTGTGGCATTCAGTCTGATCCGGAAAACGTGGAAAT	12233							
OY	68 ThrAlaGlnProTrpThrAlaProLeuProLysGlyLeuGlnCysProValProAlaSer	87							
DB	12234 GAGCAGCGGTGGTGGGAAAGCGCGTTACAAGAAAGCGGCGCAATTGCTGTCCAGGCACT	12293							
OY	88 TyrAsnAspLelPheHelleSerArgGluLileAspHisValGlyTrpValLysTrpGln	107							
DB	12294 TTTTACGATGATGTTCCGCCCATGACGATATTCGTAATTAATGTG39GCAACGTCTGGATACG	12353							
OY	108 ArgGluValLileValProLysGlyTrpSerGlnGluArgTyrLeuValAlaArgAlaGluSer	127							
DB	12354 CGCGAATCTTTTATACCGAAGAGTTGGCAGCGCAGCGATATGTCGTGGTTGCATGCG	12413							
OY	128 AlaThrHisHisGlyArgLileTyrValAlaAspAspArgLeuValAlaGlnHisValGlyGly	147							
DB	12414 GTCACTCATTTACCGCAAGATGTGTGGTCAATATACAGAAAGTATGATGAGCATCAGGCGCG	12473							
OY	148 TyrThrProPheGlnLysAlaAspValThrGlnLysLeuAlaLarGlnGlyLysPheArgLys	167							
DB	12474 TATACGCCATTTTAAGCCCAAGTGTCAAGCCCGTATGTTATTCGCGGAAAAGTACAGTATC	12533							
OY	168 ThrLileGlyValAlaAsnGlnLysLeuThrHisGlyThrLileProGlyLysLileThrThr	187							
DB	12534 ACAAGTTGTGTGAACAACGAACTGAACTGCGAAGATATCCCGCGGGAATGTGATTAAC	12593							
OY	188 GlyAsnAlaThrGlyLysValArgLileGlnThrTyrGlnHisAspPheTyrAsnTrpLacGly	207							
DB	12594 ---GACCAAAACGGCAAGAAAAGCAAGCTGTCTTCACAGATTTCTTTAACTACACCCGG	12650							
OY	208 LeuAlaArgSerLileTrpLeuTyrSerValProGlnGlnHisLileGlnAspLileThrVal	227							
DB	12651 ATCATATGCGACGGATATGCTCTACACACGCGGAACACCTGGATGGAGATATCACCGTG	12710							
OY	228 ValThrAspVal-----AspGlyAspAsnGlyLeuLileAsnTyrGlyValGluValAla	245							
DB	12711 GTGACGATATCCGCAAGACTGTAAACCAACGCGTCTGTGGACTGTGGCAAGGTG---GTGGCC	12767							

```

QY 246 AsnGlnThrThrGlyGlnIleGlnIleSerValIleAspGluAspGlyValaIleValaIle 265
Db 12768 AAT-----GGTGAATGTCAGCGCTTGAACCTGCGTATGCGGATTCACACAGCTGTTGCA 12818
QY 266 LysAlaSerGlyAlaGlnGlyThrValThrIleProSerValIlySLeuTPGInProGly 285
Db 12819 ACTGACACAGGACACAGCGGAGCTTGGCAATGGTGAATCCGCACTCTGGCAACCGGGT 12878
QY 286 AlaAlaTyrLeuTyrGlnIleGlnValAsnIleValGlySerSerGlyAspValaIleAsp 305
Db 12879 GAAGCTTATCTCTATGAACCTGTACGTC-----ACAGCCAAACCCACAGACAGCTGTGAT 12932
QY 306 ThrTyrAsnLeuAlaThrGlyValaIleValaIleValaIleValaIleValaIleVala 325
Db 12933 ATCTACCGCGTGGCGGCGTGGATCGGTCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGT 12992
QY 326 AsnGlyLysProPheTyrPheThrGlyPheGlyLysHisGluAspThrAlaValaIleArgGly 345
Db 12993 AACCAACAAACCGCTCTACTTACTGCGCTTGGCCGTCATGAGATGCGGATTTGCGCGGC 13052
QY 346 LysGlyHisAspProAlaTyrMetValHisAspPheGlnMetLysTyrPheGlyVala 365
Db 13053 AAAGGATTCGATTAACGTCGATGATGACATGACATGACATGACATGACATGACATGACATGAC 13112
QY 366 AsnSerPheArgThrSerHisTyrProTyrAlaGlnGluValaMetAspPheAlaAspArg 385
Db 13113 AACTCTACCGCTACCTGCAATTAACCTTAACCTTAACCTTAACCTTAACCTTAACCTTAACCTTA 13172
QY 386 AsnGlyLysValaValaIleAspGlnThrProAlaValaGlyLeuAsnIleAlaLeu--Met 404
Db 13173 CATGGCATCGGTGGTGGATTTGATGAACCTGCGCTTGAACCTCTCTTTAGCGAT 13232
QY 405 GlyValSerGluSerGlyAlaPro---GlnThrPheThrProAspAlaIleAsnAspLys 423
Db 13233 GGTTCGACAGCGGCAACAGCCGAAAGAACTGACGCGAAGAGGAGTCAAGCGGAGAA 13292
QY 424 ThrGlnGlnAlaHisLysGlnAlaIleArgGlyLeuIleAlaArgAspLysAsnHisAla 443
Db 13293 ACTCAGACGCGGCACTTACAGCGGATTAAGAGCTGATACCGGTGACAAAACCAACCCCA 13352
QY 444 SerValValMetTyrSerIleAlaAsnGluProAlaSerHisGluAspGlyAlaArgGlu 463
Db 13353 AGCGTGTGATGTGAGATTTGCGCAAGAACCGGATCCGTCGCAAGGTGACCGGGA 13412
QY 464 TyrPheGluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPhe 483
Db 13413 TATTTCCGCGCACTGCGGGAAGCAACCGCTAACTCGATCCGACGCTCCGATCACCTTGC 13472
QY 484 AlaAsnValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPheAspValSer 503
Db 13473 GTCATATGATGTTCTGCGACGCTCACACCGATACCATGACGATCTTTGATGTGCTG 13532
QY 504 CysIleAsnArgTyrPheGlyTyrPyrSerGlnThrGlyLysPheGlnGluAlaGluAla 523
Db 13533 TCGCTGAACCGTATTAATGCGTGTGATGTCGCAACCGGCAATTTGGAACCGGAGGAAG 13592
QY 524 AlaLeuGlnLysGluLeuHisGlyTyrPheGlnGluLysPheHisArgProIleValaMetThr 543
Db 13593 GTACGTGAAAAAGAACTTGGCGCTGCGACAGAGAAATGATCAGCCGATTAATCATCAC 13652
QY 544 GlnTyrGlyAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTyrSerGlu 563
Db 13653 GAATACCGCGTGAATACGTTAGCGGCTGCACTCAATGACACCAACATGTGAGTGA 13712
QY 564 GluPheGlnValaGlnMetLeuAspMetTyrHisArgValaPheAspArgIleGluSerMet 583
Db 13713 GAATATCAAGTGTGATGCGTGTGATGATATCACCGCGCTTTATGCGCGCAACCGCGCT 13772
QY 584 AlaGlyLysHisValaTyrAsnPheAlaAspPheGlnThrAsnLeuGlyIleIleArgVal 603
Db 13773 GTGGGTGAACAGGATATGGAATTTGCGCGATTTTGCACCTCGCAAGGCAATATTGCGCGTT 13832

```

```

QY 604 AspGlyAsnLysLysGlyValPheThrArgAspArgLysProLysAlaAlaHisSer 623
Db 13833 GCGCGTGAACAGAGGAGATCTTACCCCGGACCGCAACCGAAGTGGCGGCTTTCTG 13892
QY 624 LeuArgAlaArgTyrPheThrIleAsp 632
Db 13893 CTCGAAACAAACGTGACTGCGATGAAAC 13919
RESULT 6
ID AAA27980 standard; DNA; 4299 BP.
XX
AC AAA27980;
XX
DT 15-AUG-2000 (first entry)
XX
DE Vector MS23-pBT10-GUS nucleotide sequence.
XX
KM Ellicitor-responsive cis-element; vector MS23-pBT10-GUS; herbicide; ss;
KM chimeric promoter; pathogen infection; transgenic plant; resistance;
KM local response; genetic engineering; disease resistant crop.
XX
OS Synthetic.
XX
PN WO200029592-A2.
XX
PD 25-MAY-2000.
XX
PF 12-NOV-1999; 99NO-EP008710.
XX
PR 12-NOV-1998; 98EP-00121160.
PR 27-AUG-1999; 99EP-00116981.
XX
PA (PLAC ) MAX PLANCK GES FOERDERUNG.
XX
PT Kirsch C, Logemann E, Hahlbrock K, Rushton P, Somseich I,
XX
DR WPI; 2000-387804/33.
XX
PT Chimeric promoters mediating gene expression in plants upon pathogen
PT infection, useful for transgenic plant production comprises at least one
PT cis-acting element for direct elicitor-specific expression.
XX
PS Example; Fig 1; 73bp; English.
XX
CC This sequence represents a synthetic vector MS23-pBT10-GUS encoding
CC nucleotide sequence. The present invention relates to chimeric promoters
CC capable of mediating local gene expression in plants upon pathogen
CC infection. The chimeric promoters comprise at least one cis-element (see
CC AAA27964-A27979) capable of directing elicitor-specific expression, and a
CC minimal promoter. The chimeric promoters are useful for producing a
CC transgenic plant which has attained resistance or improved resistance
CC against a pathogen. The cis-acting element, chimeric promoter,
CC recombinant gene encoding the chimeric promoter, vector comprising the
CC chimeric promoter and a compound capable of activating the chimeric
CC promoter are useful for producing pathogen resistant plants, and for
CC identifying and/or producing compounds capable of conferring induced
CC resistance to a pathogen in a plant. A compound which specifically
CC activates or inhibits genes activated in a plant when attacked by a
CC pathogen is also useful as a plant protective agent or a herbicide. The
CC chimeric promoter provides rapid and local response to pathogen attack
CC but shows negligible activity in uninfected parts of the plants and
CC therefore is most suitable for the engineering of disease resistant crops
XX
SO Sequence 4299 BP; 1107 A; 1028 C; 1123 G; 1041 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 9,03e-139 Length: 4299
Score: 1717.00 Matches: 332
Percent Similarity: 70.02% Conservative: 93
Best Local Similarity: 54.70% Mismatches: 166
Query Match: 51.19% Indels: 16
Gaps: 9

```

```

US-10-757-093-4 (1-634) x AAA27980 (1-4299)
QY 33 HisGluGlnProLeuIleValArgProGlnArgThiSerSerArgGluLeuValAsn 52
Db 2103 CAGCGCTGAGTGGCCAGCATTCCTGCTCTGTAGAAACCCCAACCCCTGTAATCAAAA 2162
QY 53 LeuAspGlyLeuThrPheValLeu-----AlaSerGlyLeuAsnArgThrAla 69
Db 2163 CTGACGCGCTGTGGCATTCACTGATCGCCGAAAACCTGGAAATGATTCAG----- 2216
QY 70 GlnProThrAlaProLeuProGlyLeuGluCysProValProAlaSerTyAsn 89
Db 2217 CGTTGGGGGAAAGCGCGCTTACAAAGAAAGCGGCATTCGTGTGCCAGGCAATTTAAC 2276
QY 90 AspIlePheIleSerArgGluIleHisAspHisValGlyThrValTyTyGlnArgGlu 109
Db 2277 GATCAGTTCCCGCATGCGATATTCTGTAATTATGCGGCAACGCTGTGATCAGCGCGAA 2336
QY 110 ValIleValProGlyGlyTPSerGlnGluArgTyIleValArgAlaGluSerAlaThr 129
Db 2337 GTCTTATACGAAAGGTGGGAGCGCCAGCGATTCGTGTGGCTTTCGATGCGGTCACT 2396
QY 130 HisHisGlyArgIleTyValAsnAsnArgLeuValAlaGluHisValGlyGlyThr 149
Db 2397 CATTAACGCAAGTGTGCTCAATTAATCAGAAAGTGAATGAGATCAGGCGCGCTATACG 2456
QY 150 ProPheGluAlaAspValIleGlyLeuValAlaProGlyGluTyPheArgLeuThrIle 169
Db 2457 CCATTTGAAACCGATGTGACGCGCTATGTTATTCGCGGAAAGTGTACCTATCACCGTT 2516
QY 170 GlyValAsnAsnGlnLeuThrHisGluThrIleProProGlyGlyIleThrGlnAsn 189
Db 2517 TGTGTGAACAACGAACTGAATCGGACAGATATCCCGCGGAATGGATTAAC--GAC 2573
QY 190 AlaThrGlyIleAspArgIleGlnThrTyGlnHisAspPheTyAsnTyValAlaGlyLeuVal 209
Db 2574 GAAACGCGCAAGAAAGAGCTTACTTCAATATTTCTTAACTATTCGCGGAATCAT 2633
QY 210 ArgSerIleThrLeuTySerValProGlnGlnHisIleGlnAspIleThrValIleThr 229
Db 2634 CGCAGCCTAATGCTCTACACACCGCCGAAACACCTGGTGAACGATATCACGCTGTGACG 2693
QY 230 AspVal-----AspGlyAspAsnGlyLeuIleAsnTyGluValGluValAlaAsnGln 247
Db 2694 CATGTCCGCAAGACTGAACACGCGCTGTGTACGTGACGAGTG--GTGGCCAAAT-- 2747
QY 248 ThrThrGlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleValAlaTyAla 267
Db 2748 -----GTTGATGTCAAGCTTGAATCGGTGATCGGATCAACAGGTGTTCGAACGGA 2801
QY 268 SerGlyAlaGlnGlyThrValIleThrIleProSerValIleAspGluProGlnTyAlaAla 287
Db 2802 CAAGGCACTAGCGGAGCTTTCGCAAGTGTGAATCCGACCTTCGGCAACCGGGTGAAGT 2861
QY 288 TyIleuTyGlnLeuGlnValAsnIleValGlySerSerGlyAspValIleAspThrTy 307
Db 2862 TATCTCTATGAACGTGCGTC-----ACAACCAAAAGCCAGACAGAGTGTGATATCTAC 2915
QY 308 AsnLeuAlaThrGlyValArgThrValIleValAlaGlySerGlnPheLeuIleAsnGly 327
Db 2916 CCGCTTCGCGTCCGCAACCGGTCACTGAGTGAAGGCGCAACGTTCTGATTTAACAC 2975
QY 328 LysProPheTyThrThrGlyPheGlyLysHisGluAspThrAlaValArgGlyLysGly 347
Db 2976 AAACCGTTACTTACTTGTGCTTGTGCTGATGAATGACGCGACTTCGCGCAAGGA 3035
QY 348 HisAspProAlaTyMetValHisAspPheGlnLeuMetLysTrpIleGlyAlaAsnSer 367
Db 3036 TTCGATTAACGTCTGATGTGACACGACCAACCATTAATGAATGATGGGCGCAACTCC 3095
QY 368 PheArgThrSerHisTyProTyValAlaGluValMetAspPheAlaAspArgAsnGly 387
Db 368 PheArgThrSerHisTyProTyValAlaGluValMetAspPheAlaAspArgAsnGly 387

```

```

Db 3096 TACCGTACCTCGCATTAACCTTAGCGTGAAGAGATGCTCGACTGGCGCAGATGAACATGCG 3155
QY 388 IleValValIleAspGluThrProAlaValGlyLeuAsnIleAlaLeu---MetGlyVal 406
Db 3156 ATCGGTGATGATGATAAACTGCTGCTGGCTTTAACTCTTTAGGCAATGGCTTC 3215
QY 407 SerGluSerGlyAlaPro--GlnThrPheThrProAspAlaIleAsnAspTySerGln 425
Db 3216 GAACCGGCAACAGCGCAAGAAAGATGTAACAGGAAAGAGCATCAACGGGAAACTCAG 3275
QY 426 GluAlaHisLysGlnAlaIleArgGluLeuIleAlaArgAspTyAsnHisAlaSerVal 445
Db 3276 CAAGCCCACTTACAGCGATTAAGAGCTGATGCCGCTGACAAACCAACCAAGCGTG 3335
QY 446 ValMetTrpSerIleAlaAsnGluProAlaSerHisGluAspGlyAlaArgGlyTyPhe 465
Db 3336 GTGATGTGGAGTGTGCAACGAAACCGATACCGCTCCGCAAGTGCACGGAATTTTC 3395
QY 466 GluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsn 485
Db 3396 GCGCCACTGGCGGAGCAACGCGTTAACTGACCCGACGCTCCGATCACTCGCTCAAT 3455
QY 486 ValGlyThrAlaThrTyGlnLeuAspArgIleSerAspLeuPheAspValSerCySile 505
Db 3456 GTATGTTCGCAACGCTCAACCGATACCATCAGCGATCTTTGATGTGCTGCTG 3515
QY 506 AsnArgTyThrPheGlyTyTrpTySerGlnThrGlyAspLeuGluGluValAlaLeu 525
Db 3516 AACCGTTATTAACGATGATGTGCAAAAGCGGATTTGGAAACGCGCAGAAAGTACTG 3575
QY 526 GluLysGluLeuHisGlyTyProGlnGluLysPheHisArgProIleValMetThrGlyTy 545
Db 3576 GAAAGAACTTGTGGCTGCGCAGAGAAATCTGATCAGCCGATTAATCAACCGAAATC 3635
QY 546 GlyAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTrpSerGlnGluPhe 565
Db 3636 GCGGTGATACGTTAGCGCGGCTGCACTCAATGATACCAACGATGTGAGATGAAGATAT 3695
QY 566 GlnValGlnMetLeuAspMetTyHisArgValPheAspArgIleGluSerMetAlaGly 585
Db 3696 CAGTGTGCATGCTGATGATGATCAACCGCTTGTGATCGCTCAAGCGCGCTGCTGCT 3755
QY 586 GluHisValTyAsnPheAlaAspPheGlnThrAsnLeuGlyIleIleArgValAspGly 605
Db 3756 GAAACAGTATGAATTTCCGCAATTTTCGACCTTCGCAAGGCAATTTGCGCGTGGCGGT 3815
QY 606 AsnLysLysGlyValPheThrArgAspArgLysProLysAlaAlaHisSerLeuArg 625
Db 3816 AACCAAGAGGATCTTCACTCGCAACCGCAACCGAAGTCCGCGCTTTTCTGTGCA 3875
QY 626 AlaArgTrpThrSerIleAsp 632
Db 3876 AAACGCTGACCTGCAATGAAC 3896

```

RESULT 7
 AAT38397 standard; DNA; 3169 BP.
 AAT38397;
 AC AAT38397;
 XX
 DT 25-MAR-2003 (revised)
 DT 14-FEB-1997 (first entry)
 XX
 DE pTlc 99 plasmid fragment encoding recombinant antibody.
 XX
 KW Antibody; fusion protein; recombinant antibody; tumour therapy; produg;
 XX ss.
 XX
 OS Synthetic.
 XX
 PH Key Location/Qualifiers
 FT CDS 3..644
 FT /*tag= a

QY 72 TTrThrAlaProLeuProLySGlyLeuGluCySerProValProAlaSerTyraAspIle 91
 Db 5914 TGGGAAGCCGGCTTACAAAGAACCGGGCAATTGCTGTGCGACGGCATTTTAAACGATCAG 5973
 QY 92 PheIleSerArgGluIleHisAspHisValGlyTyraValTyTYrGlnArgGluValIle 111
 Db 5974 TTGGCCGATGCAATATTCGTAATATTCGCGGCAAGCTCTGTATTCAGCGCGAAGTCTTT 6033
 QY 112 ValProLySerGlyTTPSerGlnGluArgTyLeuValArgAlaGluSerAlaThrHisHis 131
 Db 6034 ATACCGAAAGTGGGCGAGCGCACCGCATGCTGTGCTTCATGCGGTTCACCTCATTTAC 6093
 QY 132 GlyArgIleTyraValAsnAsnArgLeuValAlaGluHisValGlyGlyTyTYrProPhe 151
 Db 6094 GCGAAGTGGGCTCAATATTCAGAGATGATGAGATGATGAGGCGGCTTACGCCATTT 6153
 QY 152 GluAlaAspValThrGluLeuValAlaProGlyGlyLeuPheArgLeuThrIleGlyVal 171
 Db 6154 GAAGCCGATGTCAAGCCGATGTATGTATGCGGGGAAAGTATGATGATGATGATGATG 6213
 QY 172 AsnAsnGluLeuThrHisGluThrIleProProGlyLeuSerIleThrThrGlyAsnAlaThr 191
 Db 6214 AACCAAGCACTGAACCTGGCAGACTATCCCGCGAATGGATGATACC--GACGAAAC 6270
 QY 192 GlyLeuArgIleGlnThrTyTYrGlnHisAspPheTyraAsnTyraAlaGlyLeuAlaArgSer 211
 Db 6271 GCGAAGAAAGAGAGTCTTACTTCCTTCATGATTTCTTTAATCATGCGGAATCATCCGACG 6330
 QY 212 IleTyrLeuTySerValProGlnGlnHisIleGlnAspIleThrValIleThrAspVal 231
 Db 6331 GTATGCTCTACACACCGCGAACCCTGGATGATGATGATGATGATGATGATGATGATG 6390
 QY 232 -----AspGlyAspAsnGlyLeuIleAsnTyraGluValGluValAlaAsnGlnThrThr 249
 Db 6391 GCGCAAGACTGTAAACCGCGCTGTGACTGGAGGTG--GTGGCAAT----- 6438
 QY 250 GlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleValAlaIleValAlaSerGly 269
 Db 6439 GGTGATGTACAGGTGAACCTGCGATGCGGATCAACAGGTGTGTCGATGATGATGATGATG 6498
 QY 270 AlaGlnGlyThrValThrIleProSerValIleLeuThrGlnProGlyAlaIleValIle 289
 Db 6499 ACTAGCGGCACTTGCAGAGTGAATCCGACCTCGGCAACCGGGGTGAAGGTATATCTC 6558
 QY 290 TyrGlnLeuGlnValAsnIleValGlySerSerGlyAspValValAspThrTyraAsnLeu 309
 Db 6559 TATGAACGTGCGTC-----ACAGCCAAAGCCAGACAGAGTGTATCTACCCGCTT 6612
 QY 310 AlaThrGlyValArgThrValIleValIleGlySerGlnPheLeuIleAsnGlyIlePro 329
 Db 6613 CGCGTCGCGATCCGTCAGTGAAGTGAAGGCGCAACAGTCTGTATTAACCAAAACCG 6672
 QY 330 PheTyTYrPheThrGlyPheGlyIleHisGluAspThrAlaValArgGlyIleGlyHisAsp 349
 Db 6673 TTCTACTTTACTGCTGCTTGTGCTGTGATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 6732
 QY 350 ProAlaIleMetValHisAspPheGlnLeuMetIleTyrIleGlyAlaAsnSerPheArg 369
 Db 6733 AAGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 6792
 QY 370 ThrSerHisTyTYrProTyraIleGluIleValIleMetAspPheAlaAspArgGlyIleVal 389
 Db 6793 ACCTCCCATTAACCTTAACGCTGAAGATGCTGATGATGATGATGATGATGATGATGATG 6852
 QY 390 ValIleAspGluThrProAlaValGlyLeuAsnIleAlaLeu--MetGlyValSerGlu 408
 Db 6853 GTGATTAAGTAACCTGCTGCTGCTGCTTAACTCTTTAGCATGCTTTTCGAAGCG 6912
 QY 409 SerGlyAlaPro--GlnThrPheThrProAspAlaIleAsnAspTyTYrThrGlnGluIle 427
 Db 6913 GGCACAAAGCCGAAAGAACTGTACAGCGAAGGCGAGTCAACGGGAAACTCAGCAAGCG 6972

QY 428 HisLeuGlnAlaIleArgGluLeuIleAlaArgAspTyraAsnHisAlaSerValIleMet 447
 Db 6973 CACTTACAGCCATTTAAAGCTGATGACCGGTGAACAAACCAACCAACCGGTGATG 7032
 QY 448 TrpSerIleAlaAsnGluProAlaSerHisGluAspGlyAlaArgGlyIleTyTYrPheGluPro 467
 Db 7033 TGGAGTATTGCCAACGACCGGATACCCGTCGCAAGGTGACAGGGAAATATTTCCGCCCA 7092
 QY 468 LeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGly 487
 Db 7093 CTGGCGAAGCAACCGGTAACTCGACCGCGCTGCATGATGATGATGATGATGATGATG 7152
 QY 488 ThrAlaThrTyTYrGlnLeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArg 507
 Db 7153 TTCTGGAGGCTCACACCGATACATCAGCATCTTTGATGTGTGTGCTGCGCAACCT 7212
 QY 508 TyTYrPheGlyTyTYrTySerGlnThrGlyAspLeuGluIleGluAlaAlaLeuGluIle 527
 Db 7213 TATTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 7272
 QY 528 TyTYrPheGlyTyTYrGlnGluIleAspPheHisArgProIleValMetThrGluTyTYrGlyAla 547
 Db 7273 GAACCTTCCGCTGCGAGAGAACTGATCAGCCGATTAATATCATCCGAAATTCGCGTG 7332
 QY 548 AspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTyTYrSerGluGluThrGlnVal 567
 Db 7333 GATACCTTACCGCGGCTGACATCAATGTACACCGCATGTGAGTGAAGATGATGATGATG 7392
 QY 568 GlnMetLeuAspMetTyTYrHisArgValPheAspArgIleGluSerMetAlaGlyIleHis 587
 Db 7393 GCATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 7452
 QY 588 ValTyrAspPheAlaAspPheGlnThrAsnLeuGlyIleIleArgValAspGlyIleHis 607
 Db 7453 GTATGGAATTTGCGCATTTTGGACCTGCAAGGATTAATGATGATGATGATGATGATGATG 7512
 QY 608 TyTYrGlyValPheThrArgAspArgTyProLyAlaAlaAlaHisSerLeuArgAlaArg 627
 Db 7513 AAAGGATCTTCACTCGGACCGCAACCGAAGTGGCGGCTTTCTGCTGCAAAAACGC 7572
 QY 628 TrpThrSerIleAsp 632
 Db 7573 TGGACTGGCATGAAC 7587
 RESULT 9
 AAF80285 standard; DNA; 8654 BP.
 AC AAF80285;
 XX 29-JUN-2001 (first entry)
 DT 29-JUN-2001 (first entry)
 XX Nucleotide sequence of plasmid pKRT1192.
 DE Vector; transgenesis; ttfA locus; RK2 ori; oriV; p285 protein;
 KW P382 protein; antibiotic resistance gene; nptII; transgenic plant; ss.
 XX Synthetic.
 OS
 XX Key
 FH rep_origin
 FT 1..654
 FT /tag= a
 FT /note= "ori RK2"
 FT 655..1263
 FT /tag= b
 FT /note= "ori ColE1"
 FT 1264..2603
 FT /tag= c
 FT /note= "NPT III gene coding for neomycin
 FT /note= "phosphotransferase and kanamycin resistance"
 FT 2604..4098
 FT /tag= d
 FT /note= "ttrA locus from RK2 coding for two proteins P285
 FT


```

FT      and P382 enabling an increase in the replication rate"
FT      4106..4271
FT      /*tag= e
FT      /note= "T-DNA left border"
FT      4272..4359
FT      /*tag= f
FT      /note= "nopaline synthetase terminator"
FT      4560..5556
FT      /*tag= g
FT      /note= "NPT III gene coding for neomycin
FT      phosphotransferase and kanamycin resistance"
FT      5557..5771
FT      /*tag= h
FT      /note= "nopaline synthetase promoter"
FT      5818..7717
FT      /*tag= i
FT      /note= "GUS gene coding for beta glucuronidase"
FT      7718..8447
FT      /*tag= j
FT      /note= "polyA from 35S ribosome"
FT      8474..8647
FT      /*tag= k
FT      /note= "T-DNA right border"

FT      FR2798139-A1.
XX      09-MAR-2001.
XX
XX      03-SEP-1999; 99FR-00011112.
XX      03-SEP-1999; 99FR-00011112.
XX      (MERI-) MERISTEM THERAPEUTICS SA.
XX      Gruber V, Comeau D;
XX      WPI; 2001-259847/27.
XX
XX      New vector free from non-essential elements, useful for transforming
XX      cells for protein production and for preparing transgenic plants.
XX
XX      Claim 20; Page 119-122; 180pp; French.
XX
XX      The specification describes a synthetic vector containing only those
XX      elements essential for its functionality and transgenesis of a cell
XX      (especially a plant cell). The vector consists of at most one origin of
XX      replication (ori), at most one sequence encoding a selection agent and a
XX      trfA locus encoding a protein that increases the level of plasmid
XX      replication. The vector particularly contains an RK2 ori, especially oriV
XX      from PRK2 of Escherichia coli with a broad host range, an antibiotic
XX      resistance gene (especially nptII conferring resistance to kanamycin in
XX      bacteria) and a trfA locus from PRK2 encoding the proteins P285 and P382.
XX      The vectors are used to prepare transgenic plants and transformed host
XX      cells for production of a heterologous proteins, e.g. insulin,
XX      interferon, lipase, blood proteins and anti-inflammatory agents. The
XX      present sequence represents a plasmid of the invention
XX
XX      Sequence 8654 BP; 2242 A; 2134 C; 2352 G; 1926 T; 0 U; 0 Other;
XX
XX      Alignment Scores:
XX      Pred. No.: 3.68e-138 Length: 8654
XX      Score: 1714.50 Matches: 337
XX      Percent Similarity: 69.12% Conservaive: 95
XX      Best Local Similarity: 53.92% Mismatches: 177
XX      Query Match: 51.12% Gaps: 17
XX      DB: 4 Gaps: 9
XX
XX      US-10-757-093-4 (1-634) x AAF80285 (1-8654)
XX
XX      15 ProSerLeuGlyThrProAlaAlaArgHisPheProArgGluMetCThrGlnHisGlu 34
XX      DB 5742 CCGAAGCTTGCGCGCGCTTAACACGCGTGATCCTT-AATTAAATGCACTTACGCGGT 5799

```

```

QY      35 GlnProLeuIleuIysValArgProGlnArgThrSerSerArgIleuValaIeuAsp 54
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      5800 GGTACGTCCTTAATGTTAGTCTCTAGAAACCACCCGATGAAATCAAAAAATCGAC 5859
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      55 GlyLeuTrpLysPheAlaIeu-----AlaSerGlyLeuAsnAspThrAlaGlnPro 71
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      5860 GGCCTGTGGCATTCAGTGTGATCCGAAACCTGTGAATTATATCAG-----CGTTGG 5913
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      72 TrpThrAlaProLeuProLysGlyLeuGluCysProValProAlaSerTyraAsnAspIle 91
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      5914 TGGAAAGCCGCTTACAGAAACCCGCGCAATTCGTGTCCAGGCACTTTAAGCATCAG 5973
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      92 PheIleSerArgGluIleuHisAspHisValGlyTrpValTyrrTyrglnArgGluValIle 111
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      5974 TTCGCGATGACATATTCGTATTAATTAATGCGGCAAGCTGTGATATCAGCGCAAGTCTT 6033
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      112 ValProLysGlyTrpSerGlnIleuGlyTrpLeuValArgAlaGluSerAlaThrHis 131
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      6034 ATACCGAAGTTGGGAGGCGCAGCGATCTGTCTGTCTGTTCATGCGGTCACTCATAC 6093
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      132 GlyArgIleTyrrValaAsnAsnArgLeuValAlaGluHisValGlyIleTyrrProPhe 151
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      6094 GCGAAGTGTGGTCAATTAATTCAGAAAGTATGAGATCAGGCGGCTATACGCCATT 6153
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      152 GluAlaAspValThrGluLeuValAlaProGlyIleuLysPheArgLeuThrIleGlyVal 171
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      6154 GAAGCGATGTCAACGCCGTATGTTATTCGCGGAAAGTGTAGTATCACCGTTGTGTG 6213
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      172 AsnAsnGluLeuThrHisGluThrIleProProGlyValIleThrThrGlyAsnAlaThr 191
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      6214 AACACGAATGAACTGACGACATATCCCGCGGAATGCTGATTACC--GACGAAAC 6270
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      192 GlyLysArgIleGlnThrTyrrGlnHisAspPheTyraSerTyrrAlaGlyLeuAlaArgSer 211
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      6271 GGCAGAAAGAGAGCTTAACTTCCTTCATGATTTTAACTATATCCGAAATCATCGGAGC 6330
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      212 IleTrpLeuTyrrSerValProGlnGlnHisIleGlnAspIleThrValaIleThrAspVal 231
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      6331 GTAAATGCTTACACACCGCCGAAACACCTGGTGAGCATATCACCGGTGACGCAATGTC 6390
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      232 -----AspGlyAspAsnGlyLeuIleAsnTyrrGlyValGluValaIleAsnGlnThr 249
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      6391 GCCCAAGACTGTAAACACCGCGCTGTGATCGACAGTGTG--GTGGCAAT----- 6438
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      250 GlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleValaIleAspAspGly 269
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      6439 GGTGATGTCCGCTTGAACCTGCGATGCGATCAACAGTGTGTGCACTGACCAAGGC 6498
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      270 AlaGlnGlyThrValThrIleProSerValLysLeuTrpGlnProGlyAlaIleTyrrLeu 289
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      6499 ACTAGCGGAGCTTTGCAAGTGTGAATCCGCACCTCTGGCAACCGGTGAAGTTATCTC 6558
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      290 TyrrGlnGluGlnValaAsnIleValaGlySerSerGlyAspValaIlePheTyrrAsnLeu 309
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      6559 TATGAACGTGGCTC-----ACAGCCAAAGACAGACAGATGTATATCTACCCGCTT 6612
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      310 AlaThrGlyValaArgThrValaIleValaIleGlySerGlnPheLeuIleAsnGlyPro 329
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      6613 CGCGTGGCATTCGGTACAGTGAAGGAGGAGGAGGAGGAGGAGTCTGTATTAACCAACCG 6672
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      330 PheTyrrPheThrGlyPheGlyLysHisGluAspThrAlaValaArgGlyLysGlyHisAsp 349
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      6673 TTCTACTTTATCTGCTTTGTGTGTCTCATGAAGATCGGCACTTGGCTGGCAAGATTCGAT 6732
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      350 ProAlaTyrrMetValaHisAspPheGlnLeuMetLysTrpIleGlyAlaAsnSerPheArg 369
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      6733 AACGTGATGATGTCGACGACCAACGATTAATGATGATGATGGGCACTCTACCGT 6792
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      370 ThrSerHisTyrrProTyrrAlaGluGluValaMetAspPheAlaAspArgAsnGlyIleVal 389
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      6793 ACCTCGATTAACTTAACTGATGAAGATGCTGATGGGCAAGATGAACATGATCATCTG 6852
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY      390 ValIleAspGluThrProAlaValaGlyLeuAsnIleAlaLeu--MetGlyValSerGlu 408
DB      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

```

```

Db      6853 GTGATGATGAAGAACTGCTGCTGCTTAACTCTTAGCATTGGTTGGAAGCG 6912
Qy      409 SerGlyAlaPro---GlnThrPheThrProAspAlaIleAsnIlePheThrGlnGluAla 427
Db      6913 GGCACAAACGCGGAAAGCTGATCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCG 6972
Qy      428 HisLeuGlnAlaIleArgGluLeuIleAlaArgAspIleAsnIleAlaSerValValMet 447
Db      6973 CACTTACAGCGGATTAAAGCTATAGCCGCTACAAAACCAACCAAGCGCTGATG 7032
Qy      448 TrpSerIleAlaAsnGluProAlaSerHisGluAspGlyAlaArgGluIlePheGluPro 467
Db      7033 TGGAGTATGCGCAACGAACCGGATACCGGATCCGCAAGATGACGGAATATTTGCGCCA 7092
Qy      468 LeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGly 487
Db      7093 CTGGCGGACGACGCGGTAACTGACCCGACCGGCTCGATCACCTGCGTCAATGATG 7152
Qy      488 ThrAlaThrTyrglnLeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArg 507
Db      7153 TTCTGCACGCTCACACCGATACGATCAGCGATCTTTATATGCTGCTGCTGAACCGT 7212
Qy      508 TyrPheGlyTrpTyrglnThrGlyAspLeuGluAlaIleGluAlaIleGluAla 527
Db      7213 TATTTACGATGATGATGTCACAAAGCGCGATTTGGAAACGACAGAGATCTGAAAAA 7272
Qy      528 GlnLeuHisGlyTrpGlnGluIlePheHisArgProIleValMetThrGluIleGlyAla 547
Db      7273 GAACCTTGCGCTGCGAGGAAAGCTGATCAGCCGATTCATCACCGGAATACGGGGTG 7332
Qy      548 AspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTrpSerGluGluPheGlnVal 567
Db      7333 GATACGTTACCCGCGCTGCACTCATGATACACGACATGAGTGATGAGTACACTGT 7392
Qy      568 GlnMetLeuAspMetTyrglnArgValPheAspArgIleGluSerMetAlaGlyGluHis 587
Db      7393 GCATGCGTGAATGATGATCAACCGCTCTTGATCGCTGACCGCGCTGCGTGAACAG 7452
Qy      588 ValTrpAsnPheAlaAspPheGlnThrAsnLeuGlyIleIleArgValAspGlyAsnIle 607
Db      7453 GTATGCAATTTCCGCAATTTGCGACCTCGCAAGGCAATATGCGCGGTGAACAG 7512
Qy      608 LysGlyValPheThrArgAspArgIleProIleAlaIleAlaHisSerLeuArgAlaArg 627
Db      7513 AAAGGATCTTCACTGCGACCGCAACCGAAGTCGCGCTTTCTGCGTGAACAAACGC 7572
Qy      628 TrpThrSerIleAsp 632
Db      7573 TGGACTGCGATGAAC 7587

RESULT 10
AAF80296
ID      AAF80296 standard; DNA; 8987 BP.
XX
AC      AAF80296;
XX
DT      29-JUN-2001 (first entry)
XX
DE      Nucleotide sequence of plasmid pMR11212.
XX
KW      Vector; transgenesis; trfA locus; RK2 ori; oriV; P285 protein;
XX      P382 protein; antibiotic resistance gene; nptIII; transgenic plant; ss.
OS      Synthetic.
XX
FH      Key
FT      rep_origin      Location/Qualifiers
FT                      1..654
FT                      /*tag= "a"
FT                      /note= "ori RK2"
FT      rep_origin      655..1263
FT                      /*tag= b
FT                      /note= "ori COL1"

```

```

FT      CDS
FT      1264..2603
FT      /*tag= c
FT      /note= "npt III gene coding for neomycin
FT      phosphotransferase and kanamycin resistance"
FT      CDS
FT      2604..4098
FT      /*tag= d
FT      /note= "trfA locus from RK2 coding for two proteins P285
FT      and P382 enabling an increase in the replication rate"
FT      terminator
FT      4272..4459
FT      /*tag= e
FT      /note= "nopaline synthetase terminator"
FT      CDS
FT      4575..5150
FT      /*tag= f
FT      /note= "Bar gene coding for phosphinotricin
FT      acetyltransferase and glutosinate resistance"
FT      promoter
FT      5151..5368
FT      /*tag= g
FT      /note= "nopaline synthetase promoter"
FT      promoter
FT      5369..6111
FT      /*tag= h
FT      /note= "enhanced promoter from 35S ribosome"
FT      CDS
FT      6159..8050
FT      /*tag= i
FT      /note= "GUS gene coding for beta glucuronidase"
FT      polyA_signal
FT      8051..8780
FT      /*tag= j
FT      /note= "polyA from 35S ribosome"
FT      misc_feature
FT      8807..8980
FT      /*tag= k
FT      /note= "T-DNA right border"

FR2798139-AI.
09-MAR-2001.
03-SEP-1999; 99FR-00011112.
03-SEP-1999; 99FR-00011112.
(MER-) MERISTEM THERAPEUTICS SA.
Gruber V, Comeau D;
MPI; 2001-259847/27.
New vector free from non-essential elements, useful for transforming
cells for protein production and for preparing transgenic plants.
Claim 20; Page 168-171; 180pp; French.

The specification describes a synthetic vector containing only those
elements essential for its functionality and transgenesis of a cell
(especialy a plant cell). The vector consists of at most one origin of
replication (ori), at most one sequence encoding a selection agent and a
trfA locus encoding a protein that increases the level of plasmid
replication. The vector particularly contains an RK2 ori, especially oriV
from pRK2 of Escherichia coli with a broad host range, an antibiotic
resistance gene (especially nptIII conferring resistance to kanamycin in
bacteria) and a trfA locus from pRK2 encoding the proteins P285 and P382.
The vectors are used to prepare transgenic plants and transformed host
cells for production of a heterologous proteins, e.g. insulin,
interferon, lipase, blood proteins and anti-inflammatory agents. The
present sequence represents a plasmid of the invention.

SQ      Sequence 8987 BP; 2342 A; 2205 C; 2438 G; 2002 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:      3,87e-138      Length:      8987
Score:          1714.50      Matches:      337
Percent Similarity: 69.12%      Conservative: 95
Best Local Similarity: 53.92%      Mismatches: 177
Query Match:      51.12%      Indels:      17
DB:              Gaps:      9

```

```

US-10-757-093-4 (1-634) x AAF80296 (1-8987)
QY 15 ProSerLeuGlyThrProAlaAlaArgHisPheProArgAsnGluMetThrGlnHisGlu 34
Db 6075 CCCAAGCTTGGCCGGCGGTAAACCGCGTGGATCTT--AATTAAAGCACTCAGGGGT 6132
QY 35 GlnProLeuLeuValArgProGlnArgThrSerSerArgProGluValAsnLeuAsp 54
Db 6133 GGTACAGTCCCTTATTGATACGCTGTAGAAACCCCAACCGGTGAAATCAAAAACTGAC 6192
QY 55 GlyLeuTrpIlePheAlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnPro 71
Db 6193 GGGCTGGGGCATTCAGTCTGGATCGGAAACCTGTGAATTGATCAG-----CGTTGG 6246
QY 72 TrpThrAlaProLeuProIleGlyLeuGluCybProValProAlaSerTrpAsnAspIle 91
Db 6247 TGGGAAAGCCCGCTTACAAAGAACCGGCAATTGCTGTCCAGCGAGTTTAAAGATCAG 6306
QY 92 PheIleSerArgGlnIleHisAspHisValGlyTrpValTrpTrpGlnArgValIle 111
Db 6307 TTGCGCCGATGCAAGATATTCTGTAATTATGCGGGCAACGCTGTGTATCAGCCGAAAGTCTT 6366
QY 112 ValProLeuGlyTrpSerGlnGluArgTrpLeuValArgAlaGluSerAlaThrHisHis 131
Db 6367 ATACCGAAGAGTTGGGAGCGCCAGCGGTATGCTGTGCTTCATGCGGTCACTCATTAC 6426
QY 132 GlyArgIleTrpValAsnAsnArgLeuValAlaGlnHisValGlyGlyTrpProPhe 151
Db 6427 GGCAGAAAGTGGGCTAATAATCAAGAAAGATGATGAGCATCAGGCGGCTTACCCCATTT 6486
QY 152 GlnAlaAspValThrGluLeuValAlaProGlyGluArgPheArgLeuThrIleGlyVal 171
Db 6487 GAAGCCGATGTACCGCGTATTGTTATTCGCGGAAAGGTACGATATCAGCCGTTGTG 6546
QY 172 AsnAsnGluLeuThrHisGluThrIleProGlyGlyValLeuThrArgIleAsnAlaThr 191
Db 6547 AACACAGAACTGAACCTGACAGACTATCCCGCGGAAATGGTGAATTAC--GACGAAAC 6603
QY 192 GlyLeuArgIleGlnThrTrpGlnHisAspPheTrpAsnTrpAlaGlyLeuAlaArgSer 211
Db 6604 GGCAGAAAGAGCACTCTTATCTTCATGATTTCTTAACTATGCGGAATCCATCCGAGC 6663
QY 212 IleTrpLeuTrpSerValProGlnGlnHisIleGlnAspIleThrValIleThrAspVal 231
Db 6664 GTATGCTGTACACACGCGCAACACCTGGGTGACATATCACCGTGGTACGATGTC 6723
QY 232 -----AspGlyAsnAsnGlyLeuIleAsnTrpGluValGluValAlaAsnGlnThr 249
Db 6724 GCGCAGACCTGTAAACACCGGCTGTGTGACTGCGAGGTG--GTGGCCAT----- 6771
QY 250 GlyGlnIleGlnIleSerValIleAspGlnAspGlyValIleValAlaValAsnSerGly 269
Db 6772 GGTGATGTACGACGCTGTAACGCGTGTGATGCGGATCAACAGGTGGTTGCACTGCAAGGC 6831
QY 270 AlaGlnGlyThrValThrIleProSerValIleValLeuTrpGlnProGlyAlaIleValLeu 289
Db 6832 ACTAGCGGAGCTTGGCAAGTGGTGAATCCGCACTTGGCAACCGGGTGAAGGTTATCTC 6891
QY 290 TrpGlnLeuGlnValAsnIleValIleGlySerSerGlyAspValValAspThrTrpAsnLeu 309
Db 6892 TATGAACCTGGCGTC-----ACAGCCAAAGCCAGACAGTGTATCTACCCGCTT 6945
QY 310 AlaThrGlyValArgThrValIleValAlaGlySerGlnPheLeuIleAsnGlyLeuPro 329
Db 6946 CCGCTCGCATCCGCTGAGTGGCAGTGAAGGCGCAACAGTTCCTGATTAACCAAAACCG 7005
QY 330 PheTrpPheThrArgIlePheGlyLeuHisGluAspThrAlaValArgGlyLeuGlyHisAsp 349
Db 7006 TTCTACTTAACTGGCTTGGCTGTGATGAAGATGCGCACTTGGCGCAAGGATTGAT 7065
QY 350 ProAlaTrpMetValHisAspPheGlnLeuMetLeuTrpIleGlyAlaAsnSerPheArg 369

```

```

Db 7066 AACGTGCTAGTGGTGCAGCAACGATTAATGACTGTGAGGCGCAACTCTACCGT 7125
QY 370 ThrSerHisTrpProTrpAlaGluGluValMetAspPheAlaAspArgAsnGlyIleVal 389
Db 7126 ACTTCGATTAACCTTACGCTGAAGATGCTGACTGGGCAATGAACATGGCATGTG 7185
QY 390 ValIleAspGluThrProAlaValIleGlyLeuAsnIleAlaLeu--MetGlyValSerGlu 408
Db 7186 GTGATGATGAACCTGCTGTGCTGTGCTTAACTCTCTTTAGGCATGTGTTTGAAGCG 7245
QY 409 SerGlyAlaPro--GlnThrPheThrProAspAlaIleAsnAspTrpThrGlnGluAla 427
Db 7246 GGCACACAGCCGAAAGAACTGATACAGCAAGAGCAGTCAACCGGAAACTCAGCAAGCG 7305
QY 428 HisLeuGlnAlaIleArgGluLeuIleAlaArgAspIleValAsnHisAlaSerValIleMet 447
Db 7306 CACTTACAGCGGATTAAGCTGAATGAGCGGTGACGCAAAAAACCAACCAACCGGTGATG 7365
QY 448 TrpSerIleAlaAsnGluProAlaSerHisGluAspGlyValaArgGlyTrpPheGluPro 467
Db 7366 TGGAGTATTGCCAACGAAACCGGATACCCCTCCGCAAGTGCACGGGAAATTTTCCGCCA 7425
QY 468 LeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGly 487
Db 7426 CTGCGCGAAGCAACGCGTAAATCGAACCCGACCGCTCCGATCACTCCGCTCAATGTATG 7485
QY 488 ThrAlaThrTrpGlnLeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArg 507
Db 7486 TTCTGCAGCCTCACACCGATACATCAGGATCTTGTGATGTCTGTGCTCGTAACCGT 7545
QY 508 TrpPheGlyTrpTrpSerGlnThrGlyAspLeuGluGluAlaIleAlaLeuGluVal 527
Db 7546 TATTACGATGATGATGATGCCAAACCGCGATTTGGAACGCAAGAAAGTACTGGAATAA 7605
QY 528 GlyLeuHisGlyTrpGlnGluValPheHisAspProIleValMetThrGlyTrpGlyAla 547
Db 7606 GAACCTTCTGCGCTGGCAGGAGAACTGCATCAGCCGATTAATCAACCGGAATACGGCGTG 7665
QY 548 AspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTrpSerGlnGluPheGlnVal 567
Db 7666 GATACGTTACCGCGGCTGCATCAATGTACACCGACATGTGGATGAAGATACAGTGT 7725
QY 568 GlnMetLeuAspMetTrpHisArgValPheAspArgIleGluSerMetAlaGlyGlnHis 587
Db 7726 GCATGCTGATATGATATCAACCGCTTGTGATCCGCTCAGCGCCGCTCGGTGAACAG 7785
QY 588 ValTrpAsnPheAlaAspPheGlnThrAsnLeuGlyIleIleArgValaAspGlyAsnVal 607
Db 7786 GTATGGAATTTGCCGATTTTGGACCTGCAAGGCATATTGGCGGTTGGCGGTTAACAG 7845
QY 608 IysGlyValPheThrArgAspArgIleProValAlaAlaHisSerLeuAlaGlyAlaArg 627
Db 7846 AAAGGATCTTCACTGCGCAACCGAAACGAAATCGCGGCTTCTGTGCAAAAAACGC 7905
QY 628 TrpThrSerIleAsp 632
Db 7906 TGAATCGCATGAAC 7920

```

RESULT 11
 AAF80292 standard; DNA; 9390 BP.
 AAF80292;
 29-JUN-2001 (first entry)
 Nucleotide sequence of plasmid pMRT1204.
 Vector; transgenesis; teta locus; RK2 ori; p285 protein;
 KM P382 protein; antibiotic resistance gene; mpcIII; transgenic plant; ss.
 OS Synthetic.

Feature	Location/Qualifiers
Key	1..654
rep_origin	/*tag= a /note= "ori RK2"
rep_origin	655..1263 /*tag= b /note= "ori COLE1"
CDS	1264..2603 /*tag= C /note= "NPT III gene coding for neomycin phosphotransferase and kanamycin resistance"
CDS	2604..4098 /*tag= d /note= "trfA locus from RK2 coding for two proteins P285 and P382 enabling an increase in the replication rate"
misc_feature	4106..4271 /*tag= e /note= "T-DNA left border"
terminator	4272..4559 /*tag= f /note= "nopaline synthetase terminator"
CDS	4560..5555 /*tag= g /note= "NPT III gene coding for neomycin phosphotransferase and kanamycin resistance"
promoter	5560..5771 /*tag= h /note= "nopaline synthetase promoter"
promoter	5772..6514 /*tag= i /note= "enhanced promoter from 35S ribosome"
CDS	6514..8453 /*tag= j /note= "GUS gene coding for beta glucuronidase"
polyA_signal	8454..9183 /*tag= k /note= "polyA from 35S ribosome"
misc_feature	9210..9383 /*tag= l /note= "T-DNA right border"
FR2798139-A1	
09-MAR-2001	
03-SEP-1999	99FR-00011112
03-SEP-1999	99FR-00011112
(MERI-) MERISTEM THERAPEUTICS SA.	
Gruber V, Comeau D;	
WPI; 2001-2559847/27.	
New vector free from non-essential elements, useful for transforming cells for protein production and for preparing transgenic plants.	
Claim 20; Page 149-152; 180pp; French.	

XX	Sequence	9390 BP; 2478 A; 2322 C; 2510 G; 2080 T; 0 U; 0 Other;
Alignment Scores:		
Pred. No.:	4, 1e-138	9390
Score:	1714.50	Matches: 337
Percent Similarity:	69.12%	Conservative: 95
Best Local Similarity:	53.92%	Mismatches: 177
Query Match:	51.12%	Indels: 17
DB:	4	Gaps: 9
US-10-757-093-4 (1-634) x AAF80292 (1-9390)		
QY	15 ProSerLeuGIYThrProAlaAlaArgHisPheProArgAsnGluMetThrGlnHisGlu	34
Db	6478 CCCAACTTGGCGCGCCGTTAACAGCGCGTGAATCTT--AAATTAAGTCAGACTTGGGGT	6533
QY	35 GlnProLeuIleuLeuValArgProGlnArgHisSerSerArgGluLeuValAsnLeuArg	54
Db	6536 GGTCAGTCCCTTAATGATGTCCTGTAGAAACCCCAACCCCGGAATCAAAAATCTGCAG	6595
QY	55 GIYLeuTrpIlePheAlaLeu-----AlaSerGIYLeuAsnArgThrAlaGlnPro	71
Db	6596 GGCGTGTGGCGATTAGTCTGGATGGCAAAACTGGAAATGATGACAG-----CGTGG	6649
QY	72 TrpThrAlaProLeuProIleGIYLeuGluCysProValProAlaSerTryAsnAspIle	91
Db	6650 TGGGAAGCGCCCTTCAAGAAAGCCGGCAATTGCTGTCCAGGCGAGTTTAACGATCAG	6709
QY	92 PheIleSerArgGlnIleHisAspHisValGIYTrpValTryArgHisArgGluValIle	111
Db	6710 TTCGGCGATGCAGATATTCGTAATTAATTCGGGCAAGCTCTGTATTCAGCGGAAGCTTT	6769
QY	112 ValProIleGIYTrpSerGlnArgGluArgTryLeuValArgAlaGluSerAlaThrHisHis	131
Db	6770 ATACGGAAGGTGGCGAGCGCAGCGTATCGTGGCTTCGATGGGTCATCATATAC	6829
QY	132 GIYArgIleTryValAsnAsnArgLeuValAlaGlnHisValGIYArgTryThrProPhe	151
Db	6830 GGCAAGTGTGGCTCAATATCAGAAATGTAGACATTCAGGGCGGTATACGCCATTT	6889
QY	152 GluAlaAspValThrGluLeuValAlaProGIYGluIlePheArgLeuThrIleGluVal	171
Db	6890 GAACCCGATGCACGCGCGTATGTTATTGGCGGGAAGATGTAACGTATACCGTTTGTGTG	6949
QY	172 AsnAsnGluLeuThrHisGluThrHisProProGIYLeuIleThrTrpGluAsnAlaThr	191
Db	6950 AACACGAACTGAACCTGGCAACTATTCGCGGGAATGGGATTTACC--GACGAAAC	7006
QY	192 GIYLeuArgIleGlnThrTryGlnHisAspPheTryAsnTryAlaGluLeuAlaArgSer	211
Db	7007 GGCAAGAAAAAGCAGTCTTACTTCCATGATTTCTTTAACTATGCGGGAATCCATCGCAGC	7066
QY	212 IleTrpLeuTrySerValProGlnGlnHisIleGlnAspIleThrValValIleThrAspVal	231
Db	7067 GTATGCTCTACACCAAGCCGAAACCTGGGTGAGATATTCACCGGTGGAGCGCATGTC	7126
QY	232 -----AspIleAspAsnGluLeuIleAsnTryGluValGluValAlaAsnGlnThrThr	249
Db	7127 GCGCAACACTGTAACCAACGCGCTGTGTACTGGCAGGTG---GTGGCCAAAT-----	7174
QY	250 GIYGlnIleGlnIleSerValIleAspGluAspGluAlaIleValAlaIleValAsaSerGly	269
Db	7175 GGTGATGTACGCTGTAACCTGGCTGATGCGGATCAACAGGTGGTTTCAACTGGCAAGGC	7234
QY	270 AlaGlnGIYThrValThrIleProSerValIleLeuTrpGlnProGIYAlaAlaIleTryLeu	289
Db	7235 ACTAGCGGACTTTGGCAAGTGGTAATCCGCACTTGGCAACCGGGCTAAGAGTTATCTC	7294
QY	290 TyrGlnLeuGlnValAsnIleValGIYSerSerGlyAspValValAspThrTryAsnLeu	309
Db	7295 TATGAAGCTGGCGTCTC-----ACAGCGAAAGCCAGACAGAGTGTGTATTTCTACACCGCTT	7348

Qy	310	AlaThrGlyValAlaGThrThValIleValAlaGlySerGlnPheLeuIleAsnGlyIleAsnPro	329
Db	7349	CGCGTCGGCATCCGGTCAGTGGCACTGAAGGCCAAGACGTTCTCGATTATACCACAAACCG	7408
Qy	330	PheTyRPhethrGlyPheGlyIleVehIsgIuAspThrAlaValaArgIyIleGlyIleAsp	349
Db	7409	TTCTACTTACTTGGCTTGGTGGTCATGAAGATGGCGACTTCGGTGGCAAAAGATTGCAT	7468
Qy	350	ProAlaTyRmetValHisAspPheGlnLeuMetIleTyRPIleGlyIleAsnSerPheArg	369
Db	7469	AACGTCGTGAATGGTCAGACAGCACCGCATTTAATGCACTGGATTGGGGCCACACTCTTACCGT	7528
Qy	370	ThiSerHisTyRProTyRAlaGluGluValMetAspPheAlaAspArgAsnGlyIleVal	389
Db	7529	ACCTGGCATTTACCTTACCTGAGAGAGATGCTCGATGGGCGCATGAACATGGCATTCGTG	7588
Qy	390	ValIleAspGluThrProAlaValaGlyLeuAsnIleAlaLeu--MetGlyValSerGlu	408
Db	7589	GTGATTGATGAACCTGCTGTGGCTTTAACTCTTTTAGCATTTGGCTTGGAAACCG	7648
Qy	409	SerGlyAlaPro--GlnThrPheThrProAspAlaIleAsnAspIleThiGlnIuAla	427
Db	7649	GGCAACAAGCCGGAAGAAGCTGATACAGCAAGAGGAGCTCAACGGGGAATCTAGCAACCG	7708
Qy	428	HisIleAsnAlaIleArgGluLeuIleAlaArgAspIleAsnHisAlaSerValIleMet	447
Db	7709	CACTTTACAGGCGATTAAAAGCTGATACGGCTGCAAAAACCAACCAAGGGTGTGATG	7768
Qy	448	TyrSerIleAlaAsnGluProAlaSerHisGluAspGlyAlaArgGlyTyRPhethrPro	467
Db	7769	TGGAGTATGGCCACAGAACCGGATACCCGTCGCCAAGTGCACGGGATATTCCCGCGCA	7828
Qy	468	LeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGly	487
Db	7829	CTGGCGGAAGCAACGGTAAATCGACCCGACGCGTCCGATCACTCGCTGAATGATATG	7888
Qy	488	ThrAlaThrTyRglnLeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArg	507
Db	7889	TTCTGGCAGCTCACACCGGATACCATTCAGCATCTCTTGATGTGCTGTGGCTGAACCGT	7948
Qy	508	TyrPheGlyTyRPyRSerGlnThrGlyAspLeuGluGluAlaGlyAlaIleGlnIuAsp	527
Db	7949	TATTACGATGATGATGTCMAAGCGGCGATTTGGAAACGCGACAGACGATCTGGAAATA	8008
Qy	528	GluLeuHisGlyTyRglnGluIleAspPheHisArgProIleValaMetThrGluTyRgIyAla	547
Db	8009	GAACTTTCGGCTGGCAGAGAAATCGATCAGCGGATTCATCATCAGCAATACGGCGTG	8068
Qy	548	AspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTyRSerGlnIuPheGlnVal	567
Db	8069	GATACGTTAGCCGGGCTCCACTCATATGACCCGACATGTGGAGAGTGAAGATCATAGCT	8128
Qy	568	GlnMetLeuAspMetTyRHisArgValPheAspArgIleGluSerMetAlaGlyIuHis	587
Db	8129	GCATGGCTGGATATGATACACCGCTTTGATCCGTCAGCGCGCTGTGGGTGAACAG	8188
Qy	588	ValTyRAsnPheAlaAspPheGlnThrAsnLeuGlyIleIleArgValAspGlyAsnIle	607
Db	8189	GTATGGATTTCGCCGATTTTGGCGACCTCGCAAGCATATTGGCGGTTGGGGTAAACAG	8248
Qy	608	LysGlyValPheThrArgAspArgIleProIleValaAlaAlaHisSerLeuThrArgIleArg	627
Db	8249	AAAGGAGATCTTCACTCGGACCGCAAAACGGAATCGCGGCTTTTCTCTGCAAAAAGCG	8308
Qy	628	TyrPheSerIleAsp	632
Db	8309	TGGACTGGCATGAAC	8323
RESULT 12			
AAAF80294			
ID AAF80294 standard; DNA; 9390 BP.			
XX AAF80294;			
NC			

XX	29-JUN-2001	(first entry)
DT		
XX	Nucleotide sequence of plasmid pMR1206.	
DE		
XX	Vector; transgenesis; trfA locus; RK2 ori; oriV; P285 protein;	
KW	P382 protein; antibiotic resistance gene; nptIII; transgenic plant; ss.	
XX		
OS	Synthetic.	
XX		
FH	Key	Location/Qualifiers
FT	rep_origin	1..654
FT		/tag= a
FT		/note= "ori RK2"
FT	rep_origin	655..1263
FT		/tag= b
FT		/note= "ori ColE1"
FT	CDS	1264..2603
FT		/tag= c
FT		/note= "NPT III gene coding for neomycin phosphotransferase and kanamycin resistance"
FT	CDS	2604..4098
FT		/tag= d
FT		/note= "trfA locus from RK2 coding for two proteins P28 and P382 enabling an increase in the replication rate"
FT	misc_feature	4106..4271
FT		/tag= e
FT		/note= "T-DNA left border"
FT	terminator	4272..4559
FT		/tag= f
FT		/note= "nopaline synthetase terminator"
FT	CDS	4560..5559
FT		/tag= g
FT		/note= "NPT III gene coding for neomycin phosphotransferase and kanamycin resistance"
FT	promoter	5560..5771
FT		/tag= h
FT		/note= "nopaline synthetase promoter"
FT	promoter	5772..6514
FT		/tag= i
FT		/note= "enhanced promoter from 35S ribosome"
FT	CDS	6554..8453
FT		/tag= j
FT		/note= "GUS gene coding for beta glucuronidase"
FT	polyA_signal	8454..9183
FT		/tag= k
FT		/note= "polyA from 35S ribosome"
FT	misc_feature	9210..9383
FT		/tag= l
FT		/note= "T-DNA right border"
XX		
XX	FR2798139-A1.	
XX		
PN	09-MAR-2001.	
XX		
DD		
XX	03-SEP-1999;	99FR-00011112.
PF		
XX	03-SEP-1999;	99FR-00011112.
PR		
XX	(MERI-) MERISTEM THERAPEUTICS SA.	
PA		
XX		
P1	Gruber V, Comeau D;	
XX		
DR	WPI; 2001-259847/27.	
XX		
PT	New vector free from non-essential elements, useful for transforming	
FT	cells for protein production and for preparing transgenic plants.	
XX		
PS	Claim 20; Page 158-161; 180pp; French.	
XX		
CC	The specification describes a synthetic vector containing only those	
CC	elements essential for its functionality and transgenesis of a cell	
CC	(especially a plant cell). The vector consists of at most one origin of	

CC replication (ori), at most one sequence encoding a selection agent and a
 CC triA locus encoding a protein that increases the level of plasmid
 CC replication. The vector particularly contains an RK2 ori, especially oriV
 CC from PRK2 of *Escherichia coli* with a broad host range, an antibiotic
 CC resistance gene (especially *mpcH* conferring resistance to kanamycin in
 CC bacteria) and a triA locus from PRK2 encoding the proteins p285 and p382.
 CC The vectors are used to prepare transgenic plants and transformed host
 CC cells for production of a heterologous proteins, e.g. insulin,
 CC interferon, lipase, blood proteins and anti-inflammatory agents. The
 CC present sequence represents a plasmid of the invention

XX Sequence 9390 BP; 2479 A; 2322 C; 2510 G; 2079 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	4.1e-138	Length:	9390
Score:	1714.50	Matches:	337
Percent Similarity:	69.12%	Conservative:	95
Best Local Similarity:	53.92%	Mismatches:	177
Query Match:	51.12%	Indels:	17
DB:	4	Gaps:	9

US-10-757-093-4 (1-634) x AAF80294 (1-9390)

QY 15 ProSerLeuGlyThrProAlaAlaArgHisPheProArgAsnGluMetThrGlnHisGlu 34
 DB 6478 CCAGAGCTGGCCGCGCGTTAAACAGCGGTGATCCTT--AATTAAAGTCGACTAGGGGT 6535
 QY 35 GlnProLeuLeuValArgProGlnArgThrSerSerArgGluLeuValAsnLeuAsp 54
 DB 6536 GGTACAGCCTTATGTTACCTCTCGTAGAAACCCCAACCGGTAAATCAAAAAAAGCTGAC 6595
 QY 55 GlyLeuTrpLysPheAlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnPro 71
 DB 6596 GGGCTGGCGCATTCAGTCTGAGATCGCGAAACCTGTGAATGATCAG-----CGTGG 6649
 QY 72 TrpThrAlaProLeuProLysGlyLeuGluCysProValProAlaSerGlyTrpAsnPro 91
 DB 6650 TGGGAAAGCCGTTACAGAAAGCCGGGCAATTGCTGTGCCAGCGATTAAACGATCAG 6709
 QY 92 PheIleSerArgGluLeuIleHisAspHisValGlyTrpValArgGlnArgGluVal 111
 DB 6710 TTGCGCGATGACATATTCGTATATATGCGGCAAGCTGTGATACGCCAGAGCTTT 6769
 QY 112 ValProLysGlyTrpSerGlnGluArgGlyLeuValArgAlaGluSerAlaThrHis 131
 DB 6770 ATACCAAAAGTTGGGAGCGCGATTCGTCCTGTTCCATCGCGTCACTCATTCAC 6829
 QY 132 GlyArgIleTyrValAsnAsnArgLeuValAlaGluHisValGlyGlyTrpThrProPhe 151
 DB 6830 GCGAAGGTGGGTCAATATATCGAAGATGATGACATCAGGCGGCTATACGCCATT 6889
 QY 152 GluAlaAspValThrGluLeuValAlaProGlyGluLysPheArgLeuThrIleGlyVal 171
 DB 6890 GAAGCGGATGTCACGCGTATGTTATTCGCGGAAAGTGTACGATTCACCGTTGTG 6949
 QY 172 AsnAsnGluLeuThrHisGluThrIleProProGlyLysIleThrThrGlyAsnAlaThr 191
 DB 6950 AACCAAGAACTGACCTGCGAGACTATCCGCGGGAATGTGATTAAC---GACGAAAC 7006
 QY 192 GlyLysArgIleGlnThrArgHisAspPheTyrAsnTyrAlaGlyLeuAlaArgSer 211
 DB 7007 GCGAAGAAAAGAGCTTACTTCCATGATTTCTTAACTATCCGGAATCCATGCGAAC 7066
 QY 212 IleTrpLeuTyrSerValProGlnGlnHisIleGlnAspIleThrValValThrAspVal 231
 DB 7067 GTAAATGCTTACACCGCCGAAACACTGGGTGACAGATATCACCGTGTGACGATGTC 7126
 QY 232 -----AspGlyAspAsnGlyLeuIleAsnTyrGluValAlaGluValAlaAsnGlnThr 249
 DB 7127 GCGCAAGATCTGTAACACCGCTGTGTGACTGCGAGGTG--GTGCGCAAT----- 7174
 QY 250 GlyIleIleGlnIleSerValIleAspGluAspGlyAlaIleValAlaLysAlaSerGly 269

DB 7175 GGTGATGTCAGCTTGTAAGTCTGCTGATGCGGATCAACAGGTGCTTCACTGACAGGC 7234
 QY 270 AlAGlnGlyThrValThrIleProSerValLysLeuTrpGlnProGlyValAlaTyrLeu 289
 DB 7235 ACTAAGGGGACTTTGGAGAGTGGATCCGCACCTCTGGCAACCGGGTGAAGTTATCTC 7294
 QY 290 TyrGlnLeuGlnValAsnIleValGlySerSerGlyAspValValAspThrThrAsnLeu 309
 DB 7295 TATGAACGTGGCTC-----ACAGCCAAAGCCAGACAGCTGTATATCTAACCGCTT 7348
 QY 310 AlaThrGlyValArgThrValLysValAlaGlySerGlnPheLeuIleAsnGlyLysPro 329
 DB 7349 CGCGTCGAGATCCGGTACGATGAGGAGGAGGAAACATCTCTGATTACCAAAACCG 7408
 QY 330 PheTyrPheThrGlyPheGlyLysHisGluAspThrAlaValArgGlyLysGlyHisAsp 349
 DB 7409 TTCTACTTACCTGCGCTTGTGCTCATGAAAGATGCGAGCTTGGCGAAAGATTGAT 7468
 QY 350 ProAlaTyrMetValHisAspPheGlnLeuMetLysTrpIleGlyValAsnSerPheArg 369
 DB 7469 AACGTGCTGATGCTGACAGCACACCGCATTAATGACTGGATTGGGCGCAACTTACCGT 7528
 QY 370 ThrSerHisTyrProTyrAlaGluGluValMetAspPheAlaAspArgAsnGlyIleVal 389
 DB 7529 ACCTGCAATTACCTTACCGCTGAGAGATGCTGACGTGGCAGATGAACATGCGATCGTG 7588
 QY 390 ValIleAspGluThrProAlaValGlyLeuAsnIleAlaLeu--MetGlyValSerGlu 408
 DB 7589 GTGATTGAGAAAGCTGCTGCTGCTGCTTAACTCTTTAAGGCATGGTTCCAGACG 7648
 QY 409 SerGlyAlaPro---GlnThrPheThrProAspAlaIleAsnAspLysThrGlnGluVal 427
 DB 7649 GGCACAAAGCCGAAAGAACTGTACAGCAAGAGCAAGCAACGGGAAACTCAGCAACG 7708
 QY 428 HisLysGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHisValSerValMet 447
 DB 7709 CACTTACAGCGCATTAAGAGCTGATACGCGCTGCAAAAACCAACCAAGCGTGGTGATG 7768
 QY 448 TrpSerIleAlaAsnGluProAlaSerHisGluAspGlyAlaArgGluTyrPheGluPro 467
 DB 7769 TGGAGTATTTGCCAAGAACCGGATACCGCTCGGCAAGGTGACGGGAATATTTGCGGCA 7828
 QY 468 LeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGly 487
 DB 7829 CTGGCGAAGCAACGCGTAACCTGACCGAGCGGTGACCTGCTGCTGCAATGTATG 7888
 QY 488 ThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArg 507
 DB 7889 TTCTGCGACGCTCACACCGCATCATCATCGCATCTTTGATGTGCTGCTGCAACCGT 7948
 QY 508 TyrPheGlyTyrTyrSerGlnThrGlyAspLeuGluGluAlaGluAlaIleGluLys 527
 DB 7949 TATTACGATGTATGTCCTCAAGCGGCAATTTGAAACGCGAAGAGTACTGGAATAA 8008
 QY 528 GluLeuHisGlyTyrGlnGluLysPheHisArgProIleValMetThrGluTyrGlyAla 547
 DB 8009 GAACCTTGGCTGGCGAGAAAGAACTGATCAGCGCATTAATCATCACGGAATAGCGGTG 8068
 QY 548 AspThrIleuAlaGlyLeuHisSerIleLeuGlyLeuProTrpSerGluGluPheGlnVal 567
 DB 8069 GATACGTTAGCGCGGCTGCTCATGATGATACACGACATGTGGAGTGAAGATACAGTGT 8128
 QY 568 GlnMetLeuAspMetTyrHisArgValPheAspArgIleGluSerMetAlaGlyGlnHis 587
 DB 8129 GCATGGCTGAGTATGATACACCGGCTCTTGATGTCGCTACGCCGCTGCGGTGAACAG 8188
 QY 588 ValTrpAsnPheAlaAspPheGlnThrAsnLeuGlyIleIleArgValAspGlyAsnLys 607
 DB 8189 GTATGGAATTTGCGCGATTTTGCACTCGCAAGGCAATATTGCGCGTGGCGGTACAG 8248
 QY 608 LysGlyValPheThrArgAspArgLysProLysAlaAlaAlaHisSerLeuAspGlyAlaArg 627
 DB 8249 AAAGGATCTTCACTCGGACCGCAACCGAAGTCCGGGCTTTTCTGTGCAAAAACGCG 8308


```

Db      1409 GCGTGTGCGCTTTAACCTCTTTAGGATTGTTTGAACGGGCAACAGCCGAA 1350
Qy      413 GlnThrPheThrProaspAlaIleAsnAspLysThrGlnGluAlaHisLysGlnAlaIle 432
Db      1349 GAACCTGATACGCAAGAGCAAGTCAACGGGAAACCTCAGAACGCACTTACCGAAT 1290
Qy      433 ArgGluLeuIleAlaArgAspLysAsnHisAlaSerValValMetTrpSerIleAlaAsn 452
Db      1289 AAAGACTGATACGCGGTGACAAAACCAACCAAGCTGTGATGTGGAGTATGGCCAAC 1230
Qy      453 GluProAlaSerHisGluAspGluValArgGluTyrPheGluProLeuThrAsnLeuThr 472
Db      1229 GAACCGGATACCCGCTCCGCAAGGTGACCGGATATTTCCGCCACTGGCGGAACGACG 1170
Qy      473 ArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGlyThrAlaThrTyrGln 492
Db      1169 CGTAACTGACCCGCGCGGTCCGATCACCCTCGCTCAATGTATGTTTCCGACGCTCAC 1110
Qy      493 LeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArgTyrPheGlyTyr 512
Db      1109 ACCGATACCATCGCATCTCTTGTATGTCTGTCTGCTGACCGTTATTCGATGCTAT 1050
Qy      513 SerGlnThrGlyAspLeuGluGluAlaGluAlaIleGluGluLysGluLeuHisGlyTyr 532
Db      1049 GTCCAAAGCGCGCATTTGGAAACGCGACAGAAAGTACTGGAAAAAGAACTTGGCCTCG 990
Qy      533 GlnGluLysPheHisArgProIleValMetThrGluTyrGlyValAspThrLeuAlaGly 552
Db      989 CAGGAAGAACTGATCAGCCGATATCATCACCGAAATACCGCTGTGATGACCGCGG 930
Qy      553 LeuHisSerIleLeuGluLysLeuProTyrSerGluGluPheGlnValGlnMetLeuAspMet 572
Db      929 CTGCATCTCATATGACACCGACATGTGAGTGAAGATATCATGTCATGCTGATATG 870
Qy      573 TyrHisArgValPheAspArgIleGluSerMetAlaGlyGluHisValIleThrAsnPheAla 592
Db      869 TATCACCGCGCTTTGATCGCTGACGCGCTCGCTCGTGAACAGATATGAAATTTCGCC 810
Qy      593 AspPheGlnThrAsnLeuGlyIleIleArgValAspGlyAsnLysGlyValAspThr 612
Db      809 GATTTTGCAGCCTCGCAAGGATATTTGCGGTGGGTAAACAGAAAGGATCTTCACCT 750
Qy      613 ArgAspArgLysProLysAlaIleAlaHisSerLeuArgAlaArgTyrTrpSerIleAsp 632
Db      749 CCGGACCGCAACCGAAGTGGCGGCTTTCTGCTGCAAAAACGCTGAGCTGGCATGAAAC 690

RESULT 14
ADM01287
ID      ADM01287 standard; DNA; 7510 BP.
XX
AC      ADM01287;
XX
DT      01-JUL-2004 (first entry)
XX
DE      Plasmid pUTPA-f-GUS.
XX
KW      vector; plasmid; artificial intergene region; plant;
transplastomic angiosperm; agronomic property; stress resistance;
rbcl gene; ds.
XX
OS      Synthetic.
XX
PN      WO2004029256-A2.
XX
PD      08-APR-2004.
XX
PF      15-SEP-2003; 2003WO-CU000009.
XX
PR      27-SEP-2002; 2002CU-00000208.
XX
PA      (INGG-) CENT ING GENETICA & BIOTECNOLOGIA.
XX

```

```

PI      Selman-Housein Sosa G, Aguilar Cabeza E, Gonzalez Quintero ADC;
PI      Ramos Gonzalez O;
XX
DR      MPI; 2004-316131/29.
XX
PT      DNA vector for transformation and expression in plasmids, useful e.g. for
PT      producing pharmaceutical proteins or improving agronomic properties, has
PT      gene inserted in artificial intergene region.
XX
PS      Example 2; SEQ ID NO 19; 74pp; Spanish.
XX
CC      The invention relates to a DNA vector (A) for stable transformation and
CC      expression of genes (I) in plasmids, where (I) is inserted in an
CC      artificial intergene region (AIR) formed by combining two 5'-untranslated
CC      regions (5'-UTRs) of genes that transcribe in different directions and
CC      are derived from plants of different divisions or classes. (A) are used
CC      to produce transplastomic angiosperms that have improved agronomic
CC      properties (e.g. resistance to biotic or abiotic stress) or express a
CC      very wide range of agricultural, veterinary, pharmaceutical, nutritional
CC      or industrial products, e.g. enzymes, vaccinating antigens, cytokines or
CC      immunoglobulins. Use of (A) eliminates the need for a transposon for gene
CC      insertion; inserted genes do not require promoters and terminators; and
CC      the structure of flanking sequences in (A) ensures universal
CC      applicability. Also any selection marker in (A) can be eliminated by
CC      homologous recombination. (A) provides efficient and stable expression of
CC      genes without causing any functional alterations. This sequence
CC      represents the plasmid pUTPA-f-GUS an example of the vector of the
CC      invention.
XX
SQ      Sequence 7510 BP; 2025 A; 1695 C; 1867 G; 1923 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:      3,39e-138      Length:      7510
Score:          1714.00      Matches:      331
Percent Similarity: 70.50%      Conservative: 92
Best Local Similarity: 55.17%      Mismatches: 161
Query Match:    51.10%      Indels:      16
DB:            12      Gaps:      9

US-10-757-093-4 (1-634) x ADM01287 (1-7510)
Qy      40 ValArgProGlnArgThrSerSerArgGluLeuValAsnLeuAspGlyLeuTyrLysPhe 59
Db      1924 GTACGCTCGTGTAAACCCCAACCCGCTGAACAAACAACTGCAGCCGCTGTGGCATTC 1983
Qy      60 AlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnProThrPheLysPheLeu 76
Db      1984 AGCTGGATCGCAAAACTGTGGAATTGATCAG-----CGTGGTGGAAAGCGGCTTA 2037
Qy      77 ProLysGlyLeuGluCysProValProAlaSerTyrAsnAspIlePheIleSerArgGln 96
Db      2038 CAAGAAAGCGGGCAATTGCTGTGCGGACGAGTTTAAAGATCAGTTCGCCGATGCAGAT 2097
Qy      97 IleHisAspHisValGlyTyrValTyrTyrGlnArgGluValIleValProLysGlyTyr 116
Db      2098 ATTCTGAATTAATGCGGGCAACGCTGTGATCAGCGGCAAGTCTTTATACCGAAAGGTTGG 2157
Qy      117 SerGlnGluArgTyrLeuValArgAlaGluSerAlaThrHisIleAspGlyTyrVal 136
Db      2158 GCGAGCCAGCGTATCTGCTGGGCTTTCGATGGCGTCACTCATTAACGCAAGGTGGGTC 2217
Qy      137 AsnAsnArgLeuValAlaGluHisValGlyTyrThrProPheGluAlaAspValThr 156
Db      2218 AATTAATCAGAGAGTAGTAGAGATCAGGCGGCTATACGCCATTGCAACCCGATGTCAG 2277
Qy      157 GluLeuValAlaProGlyLysPheArgLeuThrIleGlyValAsnAsnGluLeuThr 176
Db      2278 CCGTATGTATTGCGCGGAAAGTGTATACCGTATCGTGTGTAACAACGAACCTGAAC 2337
Qy      177 HisGluThrIleProProGlyLysIleThrThrGlnAlaAlaThrGlyLysArgIleGln 196
Db      2338 TGGCAGACTATCCCGCGGGAATGTGATTACC--GACGAAAAACGCAAGAAAGACAG 2394

```

OY	197	ThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaArgSerTleTyrPheLeuTyrSer	216
Db	2395	TCCTACTTCACAGATTCTTTTAACATATCCCGGAATCCATCGGAGGTATAGCTTACACC	2455
OY	217	ValProGlnGlnHisTleGlnAspTleThrValValThrAspVal-----AspGlyAsp	234
Db	2455	ACGGCGAACAACCTGGGTGGAGATATCAACGTGGAGAGCATGTGGCGCAAGACTGTACAC	2514
OY	235	AsnGlyLeuTleAsnTyrGlnValGlnValAlaAsnGlnThrThrGlyGlnTleGlnTle	254
Db	2515	CACCGCTGTGTGACTGCGCAGGTG---GTGGCCAAAT-----GGTGAATGTCAGCGTT	2562
OY	255	SerValTleAspGluAspGlyAlaTleValAlaValAsnSerGlyAlaGlnGlyThrVal	274
Db	2563	GAACCTCGGTAGTCGGGATCAACAGGTGGTGTCACTGACAGCAAGCACTAGCGGGACTTGG	2622
OY	275	ThrTleProSerValTleLeuTyrProGlnProGlyAlaAlaTyrLeuTyrGlnLeuGlnVal	294
Db	2633	CAAGTGTGGAAATCCGCACCTCTGGCAACCGGTGAAGTTATCTTATGAACGTGGCGTTC	2682
OY	295	AsnTleValGlySerSerGlyAspValValAspThrTyrAsnTleValThrGlyValArg	314
Db	2683	-----ACAGCCAAAAGCCACAAGAGTGTATATCTACCGCTTCGGGTGGCATCCGG	2738
OY	315	ThrValValValAlaGlySerGlnPheLeuTleAsnGlyValProPheThrTyrPheThrGly	334
Db	2737	TCAGTGGCAGAGGAGGCGAACAAGTCTGTATTAACAACAACCTGTACTTAACTTACCTGC	2796
OY	335	PheGlyValSerTleGlnAspThrValAlaArgGlyValSerGlyHisAspProAlaTyrMetVal	354
Db	2797	TTTGGTCTCTCATGAAAGATGCGGACTTGGGTGGCAAAAGATTGATTAACGTGCTATGGTG	2855
OY	355	HisAspPheGlnLeuMetLeuTyrTleGlyAlaAsnSerPheArgTyrSerHisTyrPro	374
Db	2857	CACAGCCACGCAATTAAATGAGCTGGATTGGGGCAACTCTCAACCTACCTCGCATTAACCT	2916
OY	375	TyrAlaGlnGluValMetAspPheAlaAspArgAsnGlyTleValValTleAspGlnThr	394
Db	2917	TACGCTGAAGAGATGCTGCTGACGGCAATGAACATGTCATCTGTGTATTTGATTAACCT	2976
OY	395	ProAlaValGlyLeuAsnTleAlaLeu---MetGlyValSerGluSerGlyAlaPro---	412
Db	2977	GCTGCTGTCGGCTTTAACTCTCTTTAGGCATGTGTTTGAAGCGGGCAACAAGCCGAAA	3033
OY	413	GlnThrPheThrProAspAlaTleAsnAspTyrThrGlnGlnValHisTleGlnAlaTle	432
Db	3037	GAACGTATCAGGAGAGGACAGCTCAACGGCGAAATCTACAGCAAGCCACTTACAGCGCAT	3096
OY	433	ArgGluLeuTleAlaArgAspTyrAsnHisValAsnValMetTyrSerTleAlaAsn	452
Db	3097	AAAGAGCTGATAGCCGCTGACAAAACAACCCAAACGCTGTATGTGAGTATTCGCAAC	3155
OY	453	GluProAlaSerHisGluAspGlyAlaArgGlyTyrPheGluProLeuThrAsnLeuThr	472
Db	3157	GAACCGGATACCCGCTCCGCAAGCTCACGGGAAATTTGCGGCCACTGGCGGAAGCAACG	3216
OY	473	ArgGlnLeuAspProThrArgProTleThrPheAlaAsnValGlyThrAlaThrTyrGln	492
Db	3217	CGTAAACTCGAACCCGACCGCTGCCATCACTCCGTCATGTATGTCTCGCAAGCTCAC	3276
OY	493	LeuAspArgTleSerAspLeuPheAspValSerCysTleAsnArgTyrPheGlyTyrTyr	512
Db	3277	ACCGATACCATCAGCGATCTCTTATGATGCTGTGCTCGTACACGTTATACGATGCTAT	3333
OY	513	SerGlnThrGlyAspLeuGlnGluValGlyValAlaLeuGlyValGlnTleHisTleGlyTyr	532
Db	3337	GTCCAAACGGCGGATTTGAAACGGCAGAGAAAGTATCTGAAGAAAGAACTTCTGGCCTGG	3396
OY	533	GlnGlnTyrPheHisArgProTleValMetThrGlnTyrGlyValAspThrTleuAlaGly	552
Db	3397	CAGAGAAACTGTCATCAGCCGATTTATCATCACCGAAATACGGGTGGATATCGTTAACCGGG	3455
OY	553	LeuHisSerTleLeuGlyLeuProTyrSerGlnGluPheGlnValGlnMetLeuAspMet	572

Db	3457	CTGCACCTCAATGTAACCGACATCTGGAGTGAAGAAGATACGTGTGCATGGCTGCATATG	3516
Oy	573	TyrHisArgValPheAspArgIleGluSerMetAlaGlyIuHisValTTPAsnPheAla	592
Db	3517	TATACACCGCGCTTGTGATCGCGTCAGCGCCCTCGCTGAGTAACAGATATGAAATTTCGCC	3576
Oy	593	AspPheGlnThrAsnLeuGlyIleIleArgValAspGlyAsnIuGlyValPheThr	612
Db	3577	GATTTTTCGCACCTTCGCAGAGCGATATTTGCGCTTGCGGTAAACAAGAGGATCTTCACT	3636
Oy	613	ArgAspArgIlySPolysAlaAlaAlaHisSerLeuArgAlaArgTTPHSerIleAsp	632
Db	3637	CGCACCGCCAAACCGAAGTCCGCGCTTTTCTGTCGCAAAACCTGACCTGCATGAAC	3696
RESULT 15			
ID	ADM01291	ADM01291 standard; DNA; 8327 BP.	
AC	ADM01291;		
XX			
DT	01-JUN-2004	(first entry)	
XX			
DE	Plasmid pVTPA-f-GUS-aada DNA sequence.		
XX			
KM	vector; plasmid; artificial intergene region; plant;		
KW	transplastomic angiosperm; agronomic property; stress resistance;		
KW	rbcl gene; ds.		
OS	Unidentified.		
XX			
PN	WO2004029256-A2.		
PD	08-APR-2004.		
PF	15-SEP-2003; 2003WO-CU000009.		
XX			
PR	27-SEP-2002; 2002CU-00000208.		
PA	(INGG-1) CENT ING GENETICA & BIOTECNOLOGIA.		
XX			
PI	Selman-Housein Sosa G, Aguilar Cabeza E, Gonzalez Quintero ADC;		
PI	Ramos Gonzalez O;		
DR	WPI; 2004-316131/29.		
XX			
PT	DNA vector for transformation and expression in plasmids; useful e.g. for		
PT	producing pharmaceutical proteins or improving agronomic properties; has		
PT	gene inserted in artificial intergene region.		
PS	Example 4; SEQ ID NO 23; 74pp; Spanish.		
XX			
CC	The invention relates to a DNA vector (A) for stable transformation and		
CC	expression of genes (I) in plasmids, where (I) is inserted in an		
CC	artificial intergene region (AIR) formed by combining two 5'-untranslated		
CC	regions (5'-UTRs) of genes that transcribe in different directions and		
CC	are derived from plants of different divisions or classes. (A) are used		
CC	to produce transplastomic angiosperms that have improved agronomic		
CC	properties (e.g. resistance to biotic or abiotic stress) or express a		
CC	very wide range of agricultural, veterinary, pharmaceutical, nutritional		
CC	or industrial products, e.g. enzymes, vaccinating antigens, cytokines or		
CC	immunoglobulins. Use of (A) eliminates the need for a transposon for gene		
CC	insertion; inserted genes do not require promoters and terminators; and		
CC	the structure of flanking sequences in (A) ensures universal		
CC	applicability. Also any selection marker in (A) can be eliminated by		
CC	homologous recombination. (A) provides efficient and stable expression of		
CC	genes without causing any functional alterations. This sequence		
CC	represents the plasmid pVTPA-f-GUS-aada, an example of the vector of the		
XX			
XX			
SO	Sequence 8327 BP; 2228 A; 1894 C; 2098 G; 2107 T; 0 U; 0 Other;		

Alignment Scores:

Pred. No.: 3,87e-138 Length: 8327
 Score: 1714.00 Matches: 331
 Percent Similarity: 70.50% Conservative: 92
 Best Local Similarity: 55.17% Mismatches: 161
 Query Match: 51.10% Indels: 16
 DB: 12 Gaps: 9

US-10-757-093-4 (1-634) x ADM01291 (1-8327)

QY 40 ValArgProGlnArgThrSerSerArgGluLeuValAsnLeuAspGlyLeuTrpIlePhe 59
 DB 1924 GAGGTCCTGTAGAAACCCACCCGGAGATCAAAAACCTCAACGCGCTGTGGCAATC 1983
 QY 60 AlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnProTrpThrIleProLeu 76
 DB 1984 AGCTCGATCGGAAACCTGTGATTCAG-----CGTGGTGGAAACCGGCTTA 2037
 QY 77 ProGlyGlyLeuGluCysProValProAlaSerTyrAsnAspIlePheIleSerArgGlu 96
 DB 2038 CAAGAAAGCCGGCAATTGCTGTGCCAGCGATTTTAAACATCATCATGTCGCGATGCAGAT 2097
 QY 97 IleHisAspHisValGlyTrpValTyrTyrGlnArgGluValIleValProGlySerTrp 116
 DB 2098 ATTCGTAATTAATCGGGCAACGTCGTGTATCAGCGCAAGCTCTTTATACGAAAGGTGG 2157
 QY 117 SerGlnGluArgTyrLeuValArgAlaGluSerAlaThrHisIleGlyArgIleTyrVal 136
 DB 2158 GCAGGCGAGGTATCGTGTGCTTCGATCGAGCGGTCACTCATACGCCAAGTGGGTG 2217
 QY 137 AsnAsnArgLeuValAlaGluHisValGlyGlyTyrThrProPheGluAlaAspValThr 156
 DB 2218 AATAAATCAGAGAGTGTAGCATCAGCGCGCTATACGCTTGAAGCGATGTCACG 2277
 QY 157 GluLeuValAlaArgGlyGluGlyPheArgLeuThrIleGlyValAsnAsnGluLeuThr 176
 DB 2278 CCTATGTTATTCGCCGAAAGTGTACGATACCGGTTGTGTGAACAACGAACTGAAC 2337
 QY 177 HisGluThrIleProProGlyLysIleThrThrGlyAsnAlaThrGlyLysArgIleGln 196
 DB 2338 TGGCAGACATATCCGCGCGGGAATGTGATTAAC---GACGAAACCGCAAGAAAGCAAG 2394
 QY 197 ThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaArgSerIleTrpLeuTyrSer 216
 DB 2395 TCTTATTCATCATATTTCTTAACTATGCGGAATCATCATCGACGCTTAATGCTCTACAC 2454
 QY 217 ValProGlnGlnHisIleGlnAspIleThrValIleThrAspVal-----AspGlyAsp 234
 DB 2455 ACCGCCAACAACCTGGGTGACGATATCACGCTGTGACGCAATGCGCAAGCTGTAC 2514
 QY 235 AsnGlyLeuIleAsnTyrGluValAlaGluAlaAsnGlnThrThrGlyGlnIleGlnIle 254
 DB 2515 CACGCGTCTGTGACGTGGCAGGTG---GTGGCCAAAT-----GGTATGTCAACGTT 2562
 QY 255 SerValIleAspGluAspGlyAlaIleValAlaLeuAlaSerGlyAlaGlnGlyThrVal 274
 DB 2563 GAACCTCGTATCGGATCAACAGGTGTGTGCACTGCAACAGGCACTAGCGGAGACTTTG 2622
 QY 275 ThrIleProSerValLysLeuTrpGlnProGlyAlaAlaIleTyrLeuTyrGlnLeuGlnVal 294
 DB 2623 CAAGTGTGATCCGACCTCTGGCAACGGGTGAAGGTATCTCATGAACGTGTCCGTC 2682
 QY 295 AsnIleValAlaGlySerSerGlyAspValValAspThrTyrAsnLeuAlaThrGlyValArg 314
 DB 2683 -----ACAGCCAAAGCCAGACAGGTGTGATATCTACCCGCTTCGCGGCACTCCG 2736
 QY 315 ThrValLysValAlaGlySerGlnPheLeuIleAsnGlyLysProPheTyrPheThrGly 334
 DB 2737 TCAGTGGCAGTGAAGGCGAACGTCCTGATTAACCAACAACCGTCTACTTACTGAGC 2796
 QY 335 PheGlyLysHisGluAspThrAlaValArgGlyLysGlyHisAspProAlaTyrMetVal 354
 DB 2797 TTGTGTCATGAATGACGCACTTGCGTGGCAAGGATTCGATTAACGTCGTATGTGTG 2856

QY 355 HisAspPheGlnLeuMetLysTrpIleGlyAlaAsnSerPheArgThrSerHisTyrPro 374
 DB 2857 CACGACCAAGCATTAATGACATGGATTTGGGCCAACTCTTACCGGTACCTCGCATTAACCT 2916
 QY 375 TyrAlaGluGluValMetAspPheAlaAspArgAsnGlyTyrLeuValIleAspGluThr 394
 DB 2917 TACGCTGAAGAGATGCTCACTGGCGAGATGAACATGCGATCGTGTGATTTGTAACCT 2976
 QY 395 ProAlaValGlyLeuAsnIleAlaLeu---MetGlyValSerGluSerGlyAlaPro--- 412
 DB 2977 GCTGCTGTGGCTTTTAACTCTCTTTAGGCATTTGTTGAAGCGGCAACAAGCCGAA 3036
 QY 413 GlnThrPheThrProAspAlaIleAsnAspLysThrGlnGluAlaHisIleGlnAlaIle 432
 DB 3037 GAACCTACAGGGAAGAGCAGTCAACCGGAAACTCAGCAAGCCGCACTTACAGCGGAT 3096
 QY 433 ArgGluLeuIleAlaArgAspLysAsnHisAlaSerValValMetTrpSerIleAlaAsn 452
 DB 3097 AAGAGCTGATAGCGGTGACAAAACCAACCAAGCGGTGTGATGTGAGTATGGCAAC 3156
 QY 453 GluProAlaSerHisGluAspGlyAlaArgGlyTyrPheGluProLeuThrAsnLeuThr 472
 DB 3157 GAACCGGATACCCGTCGCCAGAGTGCACGGAAATTTTCGCCACTGGCGGAAGCAAG 3216
 QY 473 ArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGlyThrAlaThrTyrGln 492
 DB 3217 CGTAAATCGACCCGACCGCGGTCCGATCACCTCGCTCAATGTAATGTTGCGACGCTAC 3276
 QY 493 LeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArgTyrPheGlyTyrTrp 512
 DB 3277 ACCGATACCATCAGCATCTCTTGTGTGCTGTGCTGTACCCGTTATTCAGATGTAT 3336
 QY 513 SerGlnThrGlyAspLeuGluGluAlaGluAlaLeuGluLysGluLeuHisGlyTrp 532
 DB 3337 GTCCAAAGCGGCGATTGGAACCGCAGAAAGTACTGGAAAGAACTTCGCGCTGG 3396
 QY 533 GlnGluLysPheHisArgProIleValMetThrGluTyrGlyAlaAspThrLeuAlaGly 552
 DB 3397 CAGAGAAATCTCATCAGCCGATTTATCATCACCGAATACGGCGTGTAGTCGCGGG 3456
 QY 553 LeuHisSerIleLeuGlyLeuProTrpSerGluGluPheGlnValGluMetLeuAspMet 572
 DB 3457 CTGCACTCAATGTACACGACATGTGAGTGAAGATACAGTGTGCATGCTGTGATAG 3516
 QY 573 TyrHisArgValPheAspArgIleGluSerMetAlaGlyGluHisValIleTrpAsnPheAla 592
 DB 3517 TATCACCGGCTCTTTGATGCGTCAAGCCGCTGCGGTGAACAGGTATGAAATTCGCC 3576
 QY 593 AspPheGlnThrAsnLeuGlyIleIleArgValAspGlyAsnLysLysGlyValPheThr 612
 DB 3577 GATTTTGCACCTCGCAAGGCAATATTGGCGGTGAACAAGAGGATCTTCACT 3636
 QY 613 ArgAspArgLysProLysAlaAlaAlaHisSerLeuAspGlyAlaGlyTrpThrSerIleAsp 632
 DB 3637 CCGGACCCGAAACCGAAGTGGCGGCTTTTCTGTCGCAAAAACGCTGACTGACATGAAC 3696

Search completed: March 22, 2005, 18:38:23
 Job time : 885 secs

TITLE
Powels,P.H. and van den Hondel,C.A.
Functional elements in the promoter region of the *Aspergillus nidulans* *gpdA* gene encoding glyceraldehyde-3-phosphate dehydrogenase

JOURNAL
Gene 93 (1), 101-109 (1990)

MEDLINE
91032999

PUBMED
2121607

REFERENCE
4 (bases 1 to 7599)

AUTHORS
Punt,P.

TITLE
Direct Submission

JOURNAL
Submitted (22-APR-1994) Punt P., TMO Nutrition and Food Research Institute, Lange Kleiweg 139, Rijkswijk, The Netherlands

FEATURES
Source
1..7599
/organism="synthetic construct"
/mol_type="other DNA"
/db_xref="taxon:32630"

misc_feature
1..7599
/product="Fungal B-glucuronidase expression vector pNOM102"
/evidence=experimental
/citation=[2]
1..2129
/citation=[2]
/citation=[3]
/function="promoter region of the *A.nidulans* *gpdA* gene"
/evidence=experimental
2130
/gene="gpdA"
2130
/gene="gpdA"
/citation=[3]
/function="major transcription start site"
/evidence=experimental
2178..2293
/citation=[3]
/number=1
/evidence=experimental
2302..4113
/gene="uidA"
2302..4113
/gene="uidA"
/standard_name="gus"
/citation=[2]
/codon_start=1
/evidence=experimental
/transl_table=1
/product="beta-glucuronidase"
/protein_id="CA83649.1"
/db_xref="GI:475169"
/translation="MYRPVETPTREIKLIDGLMARSLSLRENCIDQRMWESALQESRA IAVGSPNDQFADIRNAGVNVYQREVFIPKVGAGRIIVRPAVTHGCVNNO EVNEHGGYTPFADVTPIYIAGKSVRIIVCVNNELMQIIPGMVITDENKKSQSY HDPFNYAGIHRVLM.YTPNTWVDIIVTAVNEDONCHASVDMQVAVDNGSVLRAD ADQOVVATGCGISGLQVNPMLMOPGEIVELCVYAKOTECDIYPRVIGIRSVAV KGEQPLINKPFIYFGFGRHEDADLRGEGFNVLMVHDAIMDMTIGANSYRSHPIYA EEWLMDADEHGIIVIDETAVGENT.SIGIGFAGKPKRELYSEEVNGETQAHQAI AHDTITSDI.FDVICLNRIYGYVQSGDLETAKVLEKELAWQETLHPIITTEGYVD TLAGHSMTYDWMSEYQCAWMDMYHVFDRSAVVGEGVWVNFADPATSGILRVGN KKGIFTRDKRPKSAFLLQKRWTMGMFGRKPGQGGKQ"

variation
2302..2307
/gene="uidA"
/note="the region around the *arg* was mutated to generate a NcoI site (information from Clontech Inc.) verified by sequence analysis"
/citation=[2]
/replace="atgctc"
2808
/gene="uidA"
/note="mutation to remove the BamHI site in the wildtype uidA gene"
/citation=[2]

variation
2808
/gene="uidA"
/note="mutation to remove the BamHI site in the wildtype uidA gene"
/citation=[2]

misc_difference
3136
/replace="a"
/gene="uidA"
/note="copy mistake in the genbank entry; see jefferson ET AL., 1986, PNAS 83:8447-8451"

misc_difference
3560..3579
/replace="g"
/gene="uidA"
/note="sequence corrections as in Farrell and Beachy (1990) Plant Mol Biol 15:821-825"
/citation=[2]
/replace="gtgcacgggaattcgcg"
4182..4951
/standard_name="terminator region from *A.nidulans* trpC gene (containing CDS)"
/note="this region contains the 3' end of the trpC transcript including 50 codons. two major transcription ends have been identified"
/citation=[1]
/citation=[2]
/evidence=experimental
4433..4443
polYA site
misc_difference
4443..4453
/note="introduction of one base at this position generates an BglI site with was shown to be present by restriction digestion"
/citation=[1]
/replace="gccttcacgac"
4652..4660
/note="introduction of one base at this position generates an AluNI site which was shown to be present by restriction digestion"
/citation=[1]
/replace="cagcncctg"
4706..4716
polYA site
misc_difference
4740..4750
/note="introduction (actual sequence could be deletion as well) of a base at this position results in the elimination of an xmi1 site which was shown to be absent by restriction digestion"
/citation=[1]
/replace="gaatcattc"
4777..4784
/note="A NcoI site present at this position was removed by digestion, treatment with Klenow polymerase and religation"
/citation=[2]
/replace="catgcatg"
4952..7599
/product="pUC18"
/note="Vector sequences derived from pUC18 from SALI to ECoRI"
/citation=[2]

variation
4777..4784
/note="A NcoI site present at this position was removed by digestion, treatment with Klenow polymerase and religation"
/citation=[2]
/replace="catgcatg"
4952..7599
/product="pUC18"
/note="Vector sequences derived from pUC18 from SALI to ECoRI"
/citation=[2]

misc_feature
4952..7599
/product="pUC18"
/note="Vector sequences derived from pUC18 from SALI to ECoRI"
/citation=[2]

ORIGIN

Alignment Scores:

	Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	1722.00	7599	335				
Percent Similarity:	69.18%			96			
Best Local Similarity:	53.77%			172			
Query Match:	51.34%			20			
DB:	12			10			

US-10-757-093-4 (1-634) x EVPN0102 (1-7599)

QY 21 AlaAlaArgHisPhePoaArgaNgIuMeTThGhHisGluInPro----- 36

Db 2236 TCAGTTGACCTTCCCATCTTACGACGCTTACCTAAGACGTAACCCCGCTTACGACAGAC 2295

QY 37 LeuileYValaArgProGlnArgThrSerSerArgGluLeuValaInsleuArgGlyLeu 56

Db 2296 ATCACCATGATCCGCTCTGTAGAAACCCCAACCGGTGAATCAAAAAAAGTCAAGGCGCTG 2355


```

/strain="FRAJ210"
/db_xref="taxon:562"
/clone="PTRC99 DICISTR. FAB/E.C.-BETA-GLUC"
3..644
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CA003450.1"
/db_xref="GI:3713239"
/translation="MDIQMTQSPSSLSASVGDRTVITCTSSSVSYMHMYQPKAP
KLIIYSTNLASGVPSRFSGSGSDFTFTISLQPEDATAYTCHQMSYPRFGGKTK
VRIKRTVAASVFIPEPSDEOLKSGTASVCLNPFDEAKVQKMDNALOGSGNSQ
STVEDQSDKSTYSLSSTLTLISKADYKHKVVAACEVTHQGLSPVTKFNRGSC"
666..3164
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CA003451.1"
/db_xref="GI:3713240"
/translation="MEVQIQESGPGIVRPSQTLSTCTVSGFTSSGYSMMHWQRPQ
RLIEWIGYIOYSGITNPNPLSKRVMLVDTSKNQSLTSSVTADTVVYICABEDY
DYHWYFDVWGQSLVTVTVSSATKGPVPLAPCSRSTSGGTAALGCLVKQYFPEPV
TVSNMGALTSGVHPFPAVLQSSGLYSLSVTVSSSLGQTQYCNVNHKSNKVD
KVEILSGSGSWVRPVEPTREIKKLGLMAFSLDENCGIDQRMWESALQESRAIAY
PESFNQPADADIRNAGVWYOREVFIPIKMGAGORIVLPPAVTHGKYVNNQSEVM
EHQGGTPPEADVTTPVIAGKSVRIIVCNINELMQITPRGVIITDENKKKQSYFHN
FNYAGIHRSVMLYTPNTWVDITVTVHVAQDCNHSVDQVANGVSYELRBDQ
QVATVQGTSGTLQVNPMLQWQEGEYLVELCVTAKSQTRECIYPLRVGIRSVAVKE
OPLINHPFYFTGFRHEDADLRKGFMDLWVHDALMDWTGKNSYRSHYPAVKE
LMDADHGIYVIDETAAVGNLSLIGIFPAKPKELYESEVNGSTQOAHLOAKEL
IARDKNPVSVMVSIANEPTDTPQAGREKVAFLAEATRTLDPRPITCNVNFCDAT
PRTSDLPVLCNRYGNYVQSGDLETAKEVLEKELLAWOEKIHOP11ITEYGVDTLA
GHSMYTDMSERYOCAMLDMTRHVRFDRAVAVGEVWMPADPATSGILRIGANKKG
IFTRDKRPSAALFLQKRWTKMFKPKQGGKQ"

```

ORIGIN

Alignment Scores:

```

pred. No.:      2 2e-116      length:      3169
Score:          1721.50      Matches:      339
Percent Similarity: 68.17%      Conservative: 98
Best Local Similarity: 52.89%      Mismatches: 179
Query Match:      51.33%      Indels:      25
                        Gaps:      11

```

US-10-757-093-4 (1-634) x AS7359 (1-3169)

```

QY      8 SerLeuLeuSerLeuAlaAlaProSerLeuGlyThrProAla-----AlaArg 23
Db      1233 AGCGTGTGACCGCTCCAGACGCTTGCGCACCCAGACTACACCTGCACAGTGAAT 1292
QY      24 HisPheProArgangluMetThrGlnHisgluGlnProLeu-----Ile 38
Db      1293 CACAAAGCCGACGACAAACAAAGGTGACAAAGAGTTGAGCTCGGTTCTGTTGTTCC 1352
QY      39 LysValArgProGlnArgThrSerSerArgGluLeuValAsnLeuAspGlyLeuTrpLys 58
Db      1353 ATGGTACGCTCTGTGAAACCCCAACCCGTGAATCAAAAACTCCAGCGGCTGTGGGCA 1412
QY      59 PheAlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnProTrpThrAlaPro 75
Db      1413 TTCAGTGTGATCGCAAAACTGTGGAATGATCAG-----CGTTGTGGAAGGCGC 1466
QY      76 LeuProLysgluLeuGluCysProValProAlaSerTyrAsnAspIlePheIleSerArg 95
Db      1467 TTCAAAAGAAAGCCGGAATTGCTGTCCACAGCGATTTAACGATCACTTCGCGCATGCA 1526
QY      96 GlnIleHisAspHisValGlyTrpValArgTyrGlnArgGluValIleValProLysGly 115
Db      1527 GARTTTGTATTTATTCGGCGGCAACGTCGTGATACGGCGCAAGCTTTATACGAAAGGT 1586
QY      116 TrpSerGlnGluArgTyrLeuValArgAlaGluSerAlaThrHisGlyArgGlyLeuTyr 135
Db      1587 TGGGCAAGCGACGATATCTGCTGCTTTCGATCGGCTCACTCATTCGCGCAAGGTGTGG 1646

```

```

QY      136 ValAsnAspArgLeuValAlaGluHisValGlyTyrThrProPheGluAlaAspVal 155
Db      1647 GTCAATTAATCAGAAAGTATGATGAGATCAGGCGCGCTATACCGCATTTGAAGCCGATTC 1706
QY      156 ThrGluLeuValAlaProGlyGluLysPheArgLeuThrIleGlyValAsnAspGluLeu 175
Db      1707 AGCGCGTATGTTATATGCGCGGAAAGTGTACGTATCACCGTTTGTTGTGAACACCACTG 1766
QY      176 ThrHisgluThrIleProProGlyLysIleThrThrGlyAsnAlaThrGlyLysArgIle 195
Db      1767 AACTGCGACGACTATCCCGCGGAAATGGTATTACC--GACGAAACGCGCAAGAAAAG 1823
QY      196 GlnThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaArgSerIleTrpLeuTyr 215
Db      1824 CAGTCTTACTTCCATTAATTTCTTTATACATACCGCGGATCCATCGACGGTATGCTTAC 1883
QY      216 SerValProGlnGlnHisIleGlnAspIleThrValValIThrAspVal-----AspGly 233
Db      1884 AACCAAGCCGAAACACTGGGTGAGCATATCACCGGTGAGCGCATGTGCGCAACACTGT 1943
QY      234 AspAsnGlyLeuIleAsnTyrGluValGluValAlaAsnGlnThrThrGlyGlnIleGln 253
Db      1944 AACCAAGCGCTGTGTGACTGCGAGGTG--GTGGCCAAAT-----GGTATGTCAGC 1991
QY      254 IleSerValIleAspGluAspGlyValAlaIleValAlaIleValAlaSerGlyValaGlnGlyThr 273
Db      1992 GTTGAACCTCGCGATGCGGATACACAGGTGCTGCACTGCGAACAGGCACTAGCGGACT 2051
QY      274 ValIThrIleProSerValLysLeuTrpGlnProGlyAlaAlaTyrLeuTyrGlnLeuGln 293
Db      2052 TTGCAAGTGGTAATCCGACCTCTGCGAACCGGGGTGAAGTTTACTCTATCACTGATGTC 2111
QY      294 ValAsnIleValGlySerSerGlyAspValValAspThrTyrAsnLeuAlaThrGlyVal 313
Db      2112 GTC-----ACAGCCAAAGCCAGACAGAGTGTGATATCTACCCGCTCGCTCGGCAATC 2165
QY      314 ArgThrValLysValAlaGlySerGlnPheLeuIleAsnGlyLysProPheTyrPheThr 333
Db      2166 CGGTCAAGTGCAGTGAAGGCGGACACAGTTCCTGATTAACCAACACCGTTTCACTTACT 2225
QY      334 GlyPheGlyLysHisgluAspThrAlaValArgGlyLysGlyHisAspProAlaTyrMet 353
Db      2226 GCGTTTGTGTCGTCAATGAATCCGACCTGCGCAACCGGATTCGATTAACCTGCTGATG 2285
QY      354 ValHisAspPheGlnLeuMetLysTrpIleGlyValAsnSerPheArgThrSerHisTyr 373
Db      2286 GTGCAGACACGACATTAATGATGATTTGGGGCCAACTCTTACCGTACCTCGCATTTAC 2345
QY      374 ProTyrAlaGluGluValMetAspPheAlaAspArgAsnGlyIleValValIleAspGlu 393
Db      2346 CCTTACGCGTGAAGAGATGCTCGACTGCGGACATGAACATGCGATCGTGTGATGATGAA 2405
QY      394 ThrProAlaValGlyLeuAsnIleAlaLeu---MetGlyValSerGluSerGlyAlaPro 412
Db      2406 ACTGCTGCTGTGGCTTTTAACTCTCTTTCAGCATTTGTTTCGAAGCGGCAACAAGCGG 2465
QY      413 ---GlnThrPheThrProAspAlaIleAsnAspLysThrGlnGluAlaHisIleGlnAla 431
Db      2466 AAGAACTGTACGCGCAAGGACGATCAACCGGGAACCTCACCAACCGCATTTACAGGGG 2525
QY      432 IleArgGluLeuIleAlaArgAspLysAsnHisAlaSerValValMetTyrSerIleAla 451
Db      2526 ATTAAGAGCTGATACGCGGTGACAAACCAACCAAGCGTGTGATGTGAGTATTTGCC 2585
QY      452 AsnGlnProAlaSerHisgluLeuAspGlyAlaArgGluTyrPheGluProLeuThrAsnLeu 471
Db      2586 AACGAACCGGATCCGCTCGCAAGGTGACGCGGAATATTTTCGCGCACTGCGCGGAAGA 2645
QY      472 ThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGlyThrAlaThrTyr 491
Db      2646 ACCGTAACCTGACCCGACGCGCTCCGATCACCTGCGTCAATGTATGTTTCGCAAGCT 2705

```

QY 492 GlnLeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArgIlePheGlyTyr 511
 DB 2706 CACACCGATACCTTCACGCACTCTTTCATGCTGCTGCTGCAACGCTTATTCAGGATGG 2765
 QY 512 TyrSerGlnThrGlyAspLeuGluValAlaIleGluIleValGluLeuGluLeuGluLeu 531
 DB 2766 TATGTCACCAAGCGCGCATTTGGAAACGCGCAGAGAAAGGTAAGTCTGAAAAAAGAACTTTCGCC 2825
 QY 532 TrpGlnGluLeuPheIleAspProIleValMetThrGluIleValIleAspThrIleVal 551
 DB 2826 TGGCAGAGAAAGTGCATCGCGCATTTATCATACCGAATACGCGGTGATACGTTAGCC 2885
 QY 552 GlyLeuIleSerIleLeuGluLeuProIlePheSerGluIlePheGluValIleMetLeuAsp 571
 DB 2886 GGGCTCATCTCATATGACCCGACATGTGAGTGAAGATATCATGTCATGTCATGTCATG 2945
 QY 572 MetThrIleArgValPheAspArgIleGluSerMetAlaGlyIleValIleTyrAspPhe 591
 DB 2946 ATGTATCACCGCGCTTTGATCGCGTCAGCGCGCTGTCGTGTAACAGTATGGAATTC 3005
 QY 592 AlaAspPheGlnThrAsnLeuGlyIleIleArgValAspGlyAsnIleValIlePhe 611
 DB 3006 GCGGATTTTGGCGCTTCGCAAGGATATTCGCGTTCGCGTAAAGAAAGGATCTTC 3065
 QY 612 ThrArgAspArgIleProIleValAlaIleSerLeuArgIleAspGlyTyrThrSerIle 631
 DB 3066 ACTCGGACCGCAACCGAAGTGGCGGCTTTCTGCTGCAAAAAGCTGACTGGCATG 3125
 QY 632 Asp 632
 DB 3126 AAC 3128
 RESULT 3
 AR096536
 LOCUS AR096536 3169 bp DNA linear PAT 08-SEP-2000
 DEFINITION Sequence 5 from patent US 6008023.
 ACCESSION AR096536
 VERSION AR096536.1 GI:10025425
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 UNCLASSIFIED.
 REFERENCE 1 (bases 1 to 3169)
 AUTHORS Oppen M., Bouslet K. and Czech J.
 TITLE Cytoplasmic expression of antibodies, antibody fragments and
 JOURNAL Patent: US 6008023-A 5 28-DEC-1999;
 FEATURES Location/Qualifiers
 source 1..3169
 /organism="Unknown"
 /mol_type="unassigned DNA"
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,2e-116 Length: 3169
 Score: 1721.50 Matches: 339
 Percent Similarity: 68.17% Conservative: 98
 Best Local Similarity: 52.89% Mismatches: 179
 Query Match: 51.33% Indels: 25
 Gaps: 11
 US-10-757-093-4 (1-634) x AR096536 (1-3169)
 QY 8 SerLeuLeuSerLeuAlaIleProSerLeuGlyThrProIle-----AlaArg 23
 DB 1233 AGCGTGTGACCGTCCCTCAGACGCTTGGGCACTGACCTGCAAGCTGCAAT 1292
 QY 24 HisPheProArgAsnGluMetThrGlnIleGluGlnProIle-----Ile 38
 DB 1293 CACAAGCCGCAACCAAGGTGACAGAGAGTTCAGCTCGGTTCTGCTTC 1352
 QY 39 LysValArgProGlnArgThrSerSerArgGluLeuValAsnLeuAspGlyLeuTyr 58
 DB 39 LysValArgProGlnArgThrSerSerArgGluLeuValAsnLeuAspGlyLeuTyr 58

DB 1353 ATGTATCGTCTGTAGAAACCCCAACCGGTGAATCAAAAAGTCAAGCGCTGTGGCA 1412
 QY 59 PheAlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnProThrAlaPro 75
 DB 1413 TTGAGTCTGAGTCCGAAATCTGGAATGATCAG-----GTTGGTGGAAAGCGG 1466
 QY 76 LeuProIleGlyLeuGluCysProValProAlaSerTyrAsnAspIlePheIleSerArg 95
 DB 1467 TTACAGAAAGCCGGGCAATGCTGTGCCAGCGAGTTTAAAGATAGTCTGGCGATGCA 1526
 QY 96 GluIleHisAspHisValGlyTyrValIleTyrThrArgGluValIleValProIleGly 115
 DB 1527 GATATTGTAATATATCGGCAACGTCGTGTATACCGCGCAAGTCTTATACCGAAAGT 1586
 QY 116 TrpSerGlnGluArgIleLeuValArgAlaGluSerAlaThrIleHisGlyAspGlyIleTyr 135
 DB 1587 TGGGACGGCCAGGATTCGTCGCTTCGATCGCGCTCATATTCGCAAGGTGG 1646
 QY 136 ValAsnAspArgLeuValAlaGluIleValIleGlyIleTyrThrProPheGluAlaAspVal 155
 DB 1647 GTCAATATCAGAAAGTATGAGCATCAAGCGCGCTATACGCTATGGAAGCGATGTC 1706
 QY 156 ThrGluLeuValAlaProGlyGluIlePheArgLeuThrIleGlyValAsnAsnGluLeu 175
 DB 1707 ACCCGTATGTATTCGCGGAAAGTATGATACCGCTTGTGTGAACAAAGAACTG 1766
 QY 176 ThrHisGluThrIleProProGlyIleIleThrThrGlyAsnAlaThrGlyIleAspGly 195
 DB 1767 AACTGGACAGCTATCCCGCGGAATGCTATTAC--GACCAAAACGCAAGAAAG 1823
 QY 196 GlnThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaAspSerIleTyrPheTyr 215
 DB 1824 CAGTCTTACTTCATATTTCTTAACTATGCGGAGATCATCGCAAGCTATGCTTAC 1883
 QY 216 SerValProGlnGlnHisIleGlnAspIleThrValIleThrAspVal-----AspGly 233
 DB 1884 ACCACCCGCAACCTCGGTGAGCAAGATACACCGTGTGACGCGTACGCGCAAGCTGT 1943
 QY 234 AspAsnGlyLeuIleAsnTyrGluValAlaAsnGlnThrArgIleGln 253
 DB 1944 AACACGCGCTGTGTATCGGCAAGTG--GTGGCAAT-----GGTAGTCAAC 1991
 QY 254 IleSerValIleAspGluAspGlyValAlaIleValAlaIleValAspSerGlyAlaGlnIleThr 273
 DB 1992 GTTGAACGTGATATCGGATCAACAGTGTGTCACAGCAAGCACTACCGGACT 2051
 QY 274 ValThrIleProSerValIleTyrProGlnProGlyValAlaIleTyrLeuTyrGlnLeuGln 293
 DB 2052 TTGCAAGTGTGATCCGCACTCTGCAACCGGGTGAAGTATCTTATGAACGTGTC 2111
 QY 294 ValAsnIleValIleSerSerGlyAspValValAspThrTyrAsnLeuAlaThrGlyVal 313
 DB 2112 GTC-----ACAGCCAAAGCCAGACAGAGTGTATATACCGCTTCGCGTGGCATC 2165
 QY 314 ArgThrValIleValAlaGlySerGlnPheLeuIleAsnGlyIleAspProPheTyrPheThr 333
 DB 2166 CGGTCAAGTGTGATGAGGCGCAAGCTTCTTATTAACCAAAACGCTTCTTACT 2225
 QY 334 GlyPheGlyIleHisGluAspThrAlaValArgGlyIleValIleAspProAlaTyrMet 353
 DB 2226 GCGTTTGTCGTGATGAAGTTCGCACTTCGCGCAAGATTCAATACGTCGATG 2285
 QY 354 ValHisAspPheGlnLeuMetIleTyrIleGlyValAsnSerPheArgThrSerHisTyr 373
 DB 2286 GTGCAGACCAACCATTAATGATGATGAGGCGCAACCTTCATACCTGCTGCAATAC 2345
 QY 374 ProTyrAlaGluIleValMetAspPheAlaAspArgIleValIleValIleAspGlu 393
 DB 2346 CTTTACGCTGAAGATGCTCACTGCGCAAGATCAATGATCTGCTGATGATGAA 2405
 QY 394 ThrProAlaValIleLeuAsnIleAlaLeu--MetGlyValSerGluSerGlyAlaPro 412
 DB 2406 ACTGCTGCTGTGCGCTTAACTCTCTTAAAGCATGTTTCAAGCGGCAACAGCCG 2465

Percent Similarity:	68.17%	Conservative:	98
Best Local Similarity:	52.89%	Mismatches:	179
Query Match:	51.33%	Indels:	25
DB:	6	Gaps:	11
US-10-757-093-4 (1-634) x AR373248 (1-3169)			
QY	8 SerLeuSerLeuAlaAlaProSerLeuGlyThrProAla-----AlaArg	23	
DB	1233 AGCGTGGAGACCGGTCCTCCAGACAGCTTGGGACCCAGACTACCTGCAGAGTGAAT	1299	
QY	24 HisPheProAryaenGluMetThrGlnHisGluGlnProLeu-----Ile	38	
DB	1293 CACAAGCCCAACCAACCAAGGTGAGACAGAGAGTGAAGTGCCTGGTTCGTTCGTTC	1352	
QY	39 LysValAlaArgProGlnAlaArgThrSerSerArgGluLeuValaLeuLeuAspGlyLeuTrpLys	58	
DB	1353 ATGGTAGCTCTGTAGAAACCCCAACCCGTAAATCAAAAACTGCAGAGGCTGTGGGCA	1412	
QY	59 PheAlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnProTrpThrAlaPro	75	
DB	1413 TTTCAGTCTGATGTCGAAACCTGGAGATTGATCAG-----CGTTGGTGGAAAGCGCG	1466	
QY	76 LeuProLysGlyLeuGluCysProValProAlaSerTyrAsnAspIlePheIleSerArg	95	
DB	1467 TTTCACAAAGAACCGCGGCAATTGCTGTCACAGCGAGTTTAAACGATCAAGTTCGCGATGCA	1526	
QY	96 GluIleHisAspHisValGlyTyrValTyrTyrGlnArgGluValIleValProLysGly	115	
DB	1527 GATATTCGTATATTGGCGGCAAGCTCTGGTATACAGCGCAAGCTTTATACGAAAGCT	1586	
QY	116 TrpSerGlnGluArgTyrLeuValArgAlaGluSerAlaThrHisGlyArgIleTyr	135	
DB	1587 TGGCAGACGACGCTATCGCTGCTCGCTTGCATCGGTCATCTCAATTACGGCAAGGTGG	1646	
QY	136 ValAsnAspArgLeuValAlaGlnHisValGlyGlyTyrThrProPheGluAlaAspVal	155	
DB	1647 GTCAATATATCAGAAAGATGATGAGACATCAGGCGCGTATACGCCATTGTAAGCCGATGC	1706	
QY	156 ThrGluLeuValAlaProGlyGluLysPheArgLeuThrIleGlyValAsnAsnGluLeu	175	
DB	1707 ACGCGGATATTATTGCGGGAAAAAGTACGATACCGTTGTGTGAACAAAGAACTG	1766	
QY	176 ThrHisGluThrIleProProGlyLysIleThrThrLeuAsnAlaThrGlyLysArgIle	195	
DB	1767 AACTGGACAGCTATCCCGCGGGAGATGTGATTAAC--GACGAAAAACGGCAAAAAAG	1823	
QY	196 GlnThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaArgSerIleTrpLeuTyr	215	
DB	1824 CAGCTTAATCTCAATATATTTCTTTAACTAATGCGCGGATCCATGCAACGTAATCTCTAC	1883	
QY	216 SerValProGlnGlnHisIleGlnAspIleThrValValThrAspVal-----AspGly	233	
DB	1884 ACCACCGCGCAACCTGGGTGAGCAATATCAACCGTGTGACGATGTGCGGCAAGACTGT	1943	
QY	234 AspAsnIleLeuLeuLeuTyrGluValGluValAlaAsnGlnThrThrGlyGlnIleGln	253	
DB	1944 AACCAACCGCTGTGATGCTGCAGGTG--GTCGCCAAT-----GGTGAATGCAC	1991	
QY	254 IleSerValIleAspGluAspGlyAlaIleValAlaIleValAsnAspGlyAlaGlnGlyThr	273	
DB	1992 GTTGAAGCTGGTATGCGGATGCAACAGTGTGTTGCAACTGGAACAGGCACTAGCGGAGCT	2051	
QY	274 ValThrIleProSerValLysLeuTyrGlnProGlnProGlyAlaAlaIleTyrLeuTyrGlnLeuGln	293	
DB	2052 TTCAAGATGGTGAATCGCACCTCTGGCAACCGGAGTAAAGCTTATCTCTAAGACTGTGC	2111	
QY	294 ValAsnIleValGlySerSerGlyAspValValAspThrTyrAsnLeuAlaThrGlyVal	313	

/note="modified form mut3 from Aequorea victoria GFP found
in GenBank Accession number M62653"
/codon_start=1
/transl_table=11
/product="green fluorescent protein"
/protein_id="AAP31914.1"
/db_xref="GI:30523397"
/translation="MSKGEELPTGVVPIVELLDVNGHKFSVSGEGDPTATYKLT
KFTCTGKLPVPPPTVTTPTGCVGCCAARYPDHMKOHDPKSMSPGYOERTIPKQ
DGNKTRAEKFERGGDTLVNRIELKQIDFKEDGNILGHKLEFYNNSHNVITIMDKQKQ
IKNFKIRHNIEDGSVOLADHYQONTPIGDVPLLPDNLHLSYQSLSKDPNEKRDHM
VLEFVTAAGTTHGMDLYK"
/complement(7966. .8826)
/gene="bla"
/complement(7966. .8826)
/gene="bla"
/codon_start=1
/transl_table=11
/product="beta-lactamase"
/protein_id="AAP31915.1"
/db_xref="GI:30523398"
/translation="MSIQHRYVALIPPAFCIPVAPHEFTLVKQDAEDQCARGY
IBDLNSKILSFREPERPMWSTERVLCCAVISRIDAGQQLGRRIHYSQNDLVE
YSPVTEKHLTDGMTVELCSAATMSDNTAANILTTIGSPKELTFLNNMGHVRLL
DRWEPLNEAIPNDEKDTTPVAMATTLKRLTGLSELTLASPKQQLIDMEADKVAAPL
LRGALPAGWFIADKSGAGERGRGIIAALGPDPKPSRIYIVITTSQATMDERNRQIA
EIGASLIGKM"

ORIGIN

Alignment Scores:

Pred. No.: 1,42e-115 Length: 9037
Score: 1718.50 Matches: 340
Percent Similarity: 69.00% Conservative: 94
Best Local Similarity: 54.05% Mismatches: 172
Query Match: 51.24% Indels: 23
DB: 12 Gaps: 10

US-10-757-093-4 (1-634) x AY286001 (1-9037)

QY 18 G1YThrProAlaAlaArgHisPheProArgAsnGlu--MetThrGlnHisGluGlnProL 37
Db 6597 GGAAATTCCTGGCAACCCGGGGGATCCACAGAAACGCTATGACCATGATTACGCCAAGCT 6538
QY 37 eu-----1lelyValaArgProGlnArgThrSerSerArgGluLeuVala 52
Db 6537 TGGCTGCAAGTAGCTGCTTATGCTCGTCTGTAGAAACCCCAACCCCGTAATCAAAA 6478
QY 52 snleuAspGlyLeuTTPlysPheAlaLeu-----AlaSerGlyLeuAsnAspThrA 69
Db 6477 AACTCGACGGCCCTGTGGCATTCAGTCTGGATCGCAAAAACCTGGAAATTCATCAG---- 6422
QY 69 laglProThrPThrAlaProLeuProLysGlyLeuGlyCysProValProAlaSerTyrA 89
Db 6421 --CGTTGGTGGAAAGCGCTTCAAGAAAGCCGGCAATTGCTGTGGCAGCGATTTTA 6364
QY 89 snAspIlePheIleSerArgGluIleHisAspHisValGlyTyrValIYTYGlnArg 109
Db 6363 ACATCATGTTCCGGCGATGACAGATATTGTAATATAGGGCAACGCTGGTATCAAGCGG 6304
QY 109 lVAlaIleValaProLysGlyTTPSerGlnGluArgTyrLeuValaArgAlaGluSerAlaT 129
Db 6303 AAGTCTTTATACCGAAAGGTGGGAGGCGACGATTCGTCGTCGCTTTGATGACGCGTCA 6244
QY 129 hrHisHisGlyArgIleTyrValaAsnAsnArgLeuValaGluHisValaGlyGlyTyrT 149
Db 6243 CTCATTACCGCAAAAGTGGGTCAATATATACGAAAGATGAGCATACAGGGGGCTATA 6184
QY 149 hrProBheGluAlaAspValaThrGluLeuValaAlaProGlyGlyLysPheArgLeuThrI 169
Db 6183 CGGCATTTTAAGCGCATGTACAGCGCGTATGTTATTCGCCGGAATAAGTGAACGATACCG 6124
QY 169 leglValaAsnAsnGluLeuThrHisGluThrIleProProGlyLysIleThrThrGlyA 189
Db 189 ::

Db 6123 TTGTGTGNAACAACGAAGTGAAGTGGACAGCTATCCCGCGGGAAATGTGATTACC---G 6067
QY 189 snAlaThrGlyValaGlyIleGlnThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuA 209
Db 6066 ACGAAACCGCAAGAAAGAGAGCTTACTTCCATATATCTTTAACTATGCGGAAATCC 6007
QY 209 laArgSerIleTTPLeuTyrSerValProGlnGlnHisIleGlnAspIleThrValaIaT 229
Db 6006 ATGCGACGTAATGCTCTTACACCAAGCCCAACACCTGGGTGACGATACACCGTGTGA 5947
QY 229 hrAspVala-----AspGlyAspAsnGlyLeuIleAsnTyrGluValaGluValaAsnG 247
Db 5946 CGCATGTCGCGCAAGCTGTAAACCAAGCGCTGTTGACTGCGAGGTG---GTGGCCAAAT- 5891
QY 247 lnrThrThrGlyGlnIleGlnIleSerValaIleAspGluAspGlyAlaIleValaAlaLysA 267
Db 5890 -----GATGATGTCACCGTTGAACCTGCTGATGCGGATCAACAGGTGTGCAACTG 5839
QY 267 laSerGlyAlaGlnGlyThrValaThrIleProSerValaLysLeuTTPGlnProGlyAlaA 287
Db 5838 GACAAAGCACTAGCGGGAATTTGCAAGTGTGATCCGACCTCTGGCAACCGGGTGAAG 5779
QY 287 laTyrLeuTyrGlnLeuGlnValaAsnIleValaGlySerGlyAspValaIaAspThrT 307
Db 5778 GTTATCTTATGAACCTGTGCGTC-----ACAGCCAAAGCCAGACAGATGTGATATCT 5725
QY 307 yfAsnleuAlaThrGlyValaArgThrValaLysValaAlaGlySerGlnPheLeuIleAsnG 327
Db 5724 ACCGCTTCGCGTGGCATCGGTCAGTGCAGTGAAGGGCAACAGTTCGATTAAAC 5665
QY 327 lYAspProBheTyrPheThrGlyPheGlyLysHisGluAspThrAlaValaArgGlyLysG 347
Db 5664 ACAAACCGTCTTACTTATCTGAGCTTTGGTGCATGATGATCGGACCTTACGTGGCAAG 5605
QY 347 lYHisAspProAlaTyrMetValaHisAspPheGlnLeuMetLysTPIleglYAlaAsnS 367
Db 5604 GATTCGATACGTCGTGATGATGTGACAGCACGACATTAATGACTGTGATGGGGCCAACT 5545
QY 367 erPheArgThrSerHisTyrProTyrAlaGluGluValaMetAspPheAlaAspArgAsnG 387
Db 5544 CTTACCTGCTCTGCACTTACCTTACCTTACCTGTAAGAGATGCTCACTGGGGAGATGAACATG 5485
QY 387 lYIleValaIleAspGluThrProAlaValaGlyLeuAsnIleAlaLeu---MetGlyV 406
Db 5484 GCATCGTGTGATGATGATAAATGCTGCTGCTGCTTTAACTCTCTTAAGCATTTGCTT 5425
QY 406 alSerGluSerGlyAlaPro---GlnThrPheThrProAspAlaIleAsnAspLysThrG 425
Db 5424 TCGAAGCGGGCAACAAGCCGAAGAACTGTACAGCGAAGACGACATCAACCGGGAAACTTC 5365
QY 425 lngluAlaHisIleGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHisAlaSerV 445
Db 5364 AGCAAGCGCACTTACAGCGGATTAAGAGCTTAAGAGCTTACCGGTGACAAACCCACCAAGCG 5305
QY 445 alValMetTTPSerIleAlaAsnGluProAlaSerHisIleGluAspGlyAlaArgGlyTyrP 465
Db 5304 TGGTATGTGAGTATTTGCAACGAACCGGATACCGCTCCGAAAGTGCACGGGAATATT 5245
QY 465 heGluProLeuThrHisLeuThrIraArgIleLeuAspProThrArgProIleThrPheAlaA 485
Db 5244 TCGCGCCACTGGCGGAGCAAGCAAGCGCTAACTGACCCGCGGTCCGATCACCTCGCTCA 5185
QY 485 snValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPheAspValaSerCysI 505
Db 5184 ATGTATATGTTCTGCGACGCTCACACCGATACCATACGATTTCTTTGATGTGCTGTGCC 5125
QY 505 laAsnArgTyrPheGlyTyrTyrSerGlnThrGlyAspLeuGluGluAlaGluAlaAlaL 525
Db 5124 TGAACCGTTATTAACGATGTATGTCCAAAGCGGGGATTTGAAACCGCAAGAAAGTGC 5065
QY 525 euGluLysGluLeuHisIleGlyTTPGlnGluLysPheHisAspProIleValaMetThrGluT 545
Db 5064 TGGAAAAAGAACTTCTGGCTGTGGCAGAGAACTCATCAGCGATATTCATCAACCGAAT 5005

QY 545 YRG1YALAApPhrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTIPSerGluGluP 565
 DB 5004 AGGGCTGATACGTTACCGGGCTGCTCACTCAATGACACCGACATGTGAGTGAAGAGT 4945
 QY 565 heGlnValAlaInmLeuAapMetTyrHisArgValPheAspArgIleGluSerMetAlaG 585
 DB 4944 ATCAGTGTGCATGCGGATATGTATCATCCGGCTCTTTGATGCGCGTCAAGCGCGTGTG 4885
 QY 585 YGluHisValTrrPaaPheAlaAspPheGlnThrAsnLeuGlyIleIleArgValAspG 605
 DB 4884 GTGAACAGGTATGAGATTTTCGCGATTTTCGCGACTTCGCAAGGCATATTGCGCGTGGCG 4825
 QY 605 YAsnLeuYsgGlyValPheThrArgAspArgPyrProLysAlaAlaHisSerLeuA 625
 DB 4824 GTTACCAAGAAAGGAGCTTCTCATCTCGACCGCAACCGAAGTGGCGCTTTCTGCTGC 4765
 QY 625 rGAlaArgTrrPrrSerIleAsp 632
 DB 4764 AAAAAGCTGAGCTGCGCATGAC 4742
 RESULT 6
 AY237647 10313 bp DNA circular SYN 22-MAR-2003
 LOCUS Cloning vector pHRGFPUS, complete sequence.
 DEFINITION AY237647
 ACCESSION AY237647
 VERSION AY237647.1 GI:29150587
 KEYWORDS Cloning vector pHRGFPUS
 SOURCE Cloning vector pHRGFPUS
 ORGANISM other sequences; artificial sequences; vectors.
 REFERENCE 1 (sites)
 AUTHORS Quandt, V. and Hynes, M.F.
 TITLE Versatile suicide vectors which allow direct selection for gene replacement in gram-negative bacteria
 JOURNAL Gene 127 (1), 15-21 (1993)
 MEDLINE 93252274
 PUBMED 8486283
 REFERENCE 2 (sites)
 AUTHORS Metcalf, W.W. and Wanner, B.L.
 TITLE Construction of new beta-glucuronidase cassettes for making transcriptional fusions and their use with new methods for allele replacement
 JOURNAL Gene 129 (1), 17-25 (1993)
 MEDLINE 93281120
 PUBMED 8335256
 REFERENCE 3 (bases 1 to 2642; 6610 to 7326)
 AUTHORS Ouahran, B., Portet, F., Teyssier, J., Liautaud, J.P. and Kohler, S.
 TITLE pBBR-GFP: a broad-host-range vector for prokaryotic promoter studies
 JOURNAL Biotechniques 26 (4), 620-622 (1999)
 MEDLINE 99275364
 PUBMED 10343896
 REFERENCE 4 (bases 1 to 10313)
 AUTHORS Ramos, H.J.O., Soares-Ramos, J.R.L., Souza, F.M., Soares-Ramos, J.R.L., Hungria, M. and Pedrosa, F.O.
 TITLE Monitoring Azospirillum-bacterial interactions using the gfp and gusA genes constitutively expressed from a new broad-host range vector
 JOURNAL J. Biotechnol. 97 (3), 243-252 (2002)
 MEDLINE 22079789
 PUBMED 12084480
 REFERENCE 5 (bases 1 to 10313)
 AUTHORS Ramos, H.J.O., Soares-Ramos, J.R.L., Souza, F.M. and Pedrosa, F.O.
 TITLE Direct Submission
 JOURNAL Submitted (14-FEB-2003) Department of Biochemistry and Molecular Biology, Universidade Federal do Paraná - UPR, Centro Politécnico, Curitiba, PR 81531-990, Brazil
 COMMENT Mobilization (mob) and replication (rep) regions are derived from the broad-host-range vector pBBR1 isolated from Bordetella bronchiseptica. The expression of uidA and gfp genes is under control of gentamycin resistance aacC1 promoter. The pgen promoter and part of the gentamycin gene (aacC1) were isolated from pUQ200KS

FEATURES as a BglII/BamHI fragment.
 source Location/Qualifiers
 1..10313
 /organism="Cloning Vector pHRGFPUS"
 /mol_type="genomic DNA"
 /db_xref="taxon:2283838"
 /note="Broad-host-range vector expressing GFP"
 1..2642
 /note="mob and rep regions from broad-host-range vector pBBR1-GFP as deposited in GenBank Accession number AF110459"
 gene complement (2982..3776)
 /gene="aph"
 /note="NPTII; derived from vector pMW6"
 complement (2982..3776)
 /gene="aph"
 /codon_start=1
 /transl_table=1
 /product="aminoglycoside-3'-O-phosphotransferase"
 /protein_id="AA063174.1"
 /db_xref="GI:29150589"
 /translation="MIBDGLHAGSPAMVERLFGYDAAQOTIGSDAAVPLSLAQR
 PVLFFKTLDSALNELQDEARLSWLTATGVCALVDVITAGSDMLGSRPGD
 LSSHLAPAEKYSIMADAMKRLHTLDPATCPDPAQKHRLERARTMEAGLVQDQDL
 EHOGLAPAELEFARLAKAMPDGEDLVITGDACTLPIMVNGRFSGFIDGRLGVADRY
 ODIALATRDIAEELGEMADRFVLVYGTAAPDSQSLAFYRLDERR"
 complement (3865..3914)
 /function="neomycin and kanamycin resistance aph (NPTII)
 gene promoter region"
 complement (4705..6517)
 /gene="uidA"
 /note="promoterless uidA gene from pMW6"
 complement (4706..6517)
 /gene="uidA"
 /codon_start=1
 /transl_table=1
 /product="beta-glucuronidase"
 /protein_id="AA063175.1"
 /db_xref="GI:29150589"
 /translation="WRRPVETPTREIKLDDGMAASLDRENGCIDORMMESALQESRA
 IAWPSFNDQFADADIRNAGVWVQREFFITKAGACQIVLFDVATHTGKAWWQ
 EAWHOGQGTLPREADVTPVILAGKSVIRIVCNVNEIMQTIIPGAVITDENKQSY
 FHDFFNYAGIHSVNLVTPPTWVDIITVTHVADCNHASVDMQVANGDVSVELRD
 ADOOVVATGCGTSGTLQVNNPILMOPBEGVLELCTVATGSCDCEIDPLRAGISVAV
 KKEQPLINRKPPTFGFEGHEDADLRKGFNDVILWHDALMDWIGANSVRSHPYA
 EEMLWADSHGIVIDETAVGPNLSLIGFPAKPKSLYSBEAVNGSTQAHLOAI
 KELIARDKNHPSVVMVSIANEPTDRPQGRFVAPLAEXTRLDPTPTTCVNVWCD
 AHTDIISDLFDVLCNRYGMYVQSGDLSTAKVLEKELAAOEKLP1IITEYGD
 TLAGHSWYTDWMSSEYOCAMDMDYHRVDRVSAVVEQVWNNFADPATSGILRVGNN
 KKGITRDRKPKSAAFLQKRWTKGNPFGRKPGQGGKQ"
 complement (6610..7326)
 /gene="gfp"
 /note="promoterless gfp gene from pBBR1-GFP; modified form
 mut3 from Aequorea victoria GFP as deposited in GenBank
 Accession number M62653"
 complement (6610..7326)
 /gene="gfp"
 /note="modified form mut3 from Aequorea victoria GFP as
 deposited in GenBank Accession number M62653"
 /codon_start=1
 /transl_table=1
 /product="green fluorescent protein"
 /protein_id="AA063176.1"
 /db_xref="GI:29150590"
 /translation="MSKBEFLTGAVPILVELDGVNCHKFSVSGEGDATYKTLT
 KILCTTGKLPVMPPLVTFFGYGVQCFAPYDPHMKQHDPSKAMPBGVQESTTIPKD
 DQNVYKRAVKFEGDVLNRIELKQIDFEDNLIQHKLELYVNSHNVIMDKQNG
 IKVNPRIHNIEDGVSOLADHYQNTPIGDSVLLPDNHYLSTQSLSDPBEKRDHM
 VLEFVTAAGITHGMDLYK"
 7363..7860
 /note="from pUQ200KS"
 complement (7865..7893)
 /note="35 and -10 promoter regions of the gentamycin

misc_feature
 7906..8633
 /note="from PJ0200KS"
 complement(9242..10102)
 /gene="bla"
 /note="derived from pBBR1-GFP"
 complement(9242..10102)
 /gene="bla"
 /codon_start=1
 /transl_table=11
 /product="beta-lactamase"
 /protein_id="AA063177.1"
 /db_xref="GI:29150591"
 /translation="MSIQHFRVALIPFPAFCLPVPAHPETLVKVDADQLGARVY
 IELDASGKILESFRPEERPMWSTFKVLGAVLSRIDAGSGLGRRIHYSDNDLVE
 YSPVPEKHLTDGKTVRELCSAATMSDNDRAAILTTGSGKETLFIANNMGNVRL
 DREPELNEAIPIVDERDITTPVAMATTKLKLGLIELLASRQQLIDMWDADNVAPRL
 LNSALPAGWPIADKSGAGRSKGIIALGLPDKPSRIIVITYTTSQATMDERNQIA
 EIGASLIKIM"

ORIGIN

Alignment Scores:
 Pred. No.: 1.69e-115 Length: 10313
 Score: 1718.50 Matches: 340
 Percent Similarity: 69.00% Conservative: 94
 Best Local Similarity: 54.05% Mismatches: 172
 Query Match: 51.24% Indels: 23
 DB: 12 Gaps: 10

US-10-757-093-4 (1-634) x AY237647 (1-10313)

QY 18 GLYThrProAlaIalaIrgHisPheProArgAsnGlu--MetThrGlnHisGluGlnProI 37
 Db 6597 GGATTCCTGCAACCGCGGGGATCCACAGAAACAGCTAGACCATGATTAACCCAGCT 6538
 QY 37 eu-----11eLyValArgProGlnArgThrSerArgGluLeuVala 52
 Db 6537 TGGCTCAGGTCACTCCCTTATGTCGTCCTAGAAACCCCAACCCGTAATCAAAA 6478
 QY 52 snLeuArgGlyLeuThrPheAlaLeu-----AlaSerGlyLeuAsnAspTha 69
 Db 6477 AACTCAGCGCGCTGTGGCATTTCAGTCTGATGCCAATACTGTGAATGATGACAG-- 6422
 QY 69 lagInProThrAlaProLeuProLeuGlyLeuGlyCysProValProAlaSerTyra 89
 Db 6421 --CGTTGGGGGAAACCGCTTACAGAAACCGGGCAATGCTGTGCGCAGGAGTTTA 6364
 QY 89 snAspIlePheIleSerArgGluIleHisAspHisValGlyTyraIleTyrglnArg 109
 Db 6363 ACCATCAGTTCGCCGAGATATTCGTAATATGCGGGCAACGTCGTGATCAACGCG 6304
 QY 109 luValIleValProLeuGlyTyrSerGlnGluArgTyrLeuValArgAlaGluSerAlaT 129
 Db 6303 AAGTCTTTATACCGGAAGGTGGGCAAGCGCATTCGTCGCGTTCGATGCCGCTCA 6244
 QY 129 hrHisHisGlyArgIleTyrValAsnAspArgLeuValAlaGlnHisValGlyGlyTyrT 149
 Db 6243 CTCATTACGGCAAGGTGGGTCAATATACAGAAATGATGACATCAGGCGCGCTTAA 6184
 QY 149 hrProPheGluAlaAspValThrGlyLeuValAlaProGlyGlyLeuPheArgLeuThrI 169
 Db 6183 CGCATTTAAGCGCATGTCACCCCGTATGTATTCCCGGAAATGTATCATCAACCG 6124
 QY 169 lsgLyValAsnAspGlyLeuThrHisGlyThrIleProProGlyTyrIleThrThrGlyA 189
 Db 6123 TTGTGTGAACAAGAACTGAACCTAGCTATCCCGCGGAAATGATGATTACC--G 6067
 QY 189 snAlaThrGlyLysArgIleGlnThrTyrglnHisAspPheTyraAsnTyraAlaGlyLeuA 209
 Db 6066 ACGAAACGGCAAGAAAGAGCTCTTACTTCATATTTCTTTAATCTATGCCGGAATCC 6007
 QY 209 lArgSerIleTyrLeuTyrSerValProGlnGlnHisIleGlnAspIleThrValValT 229
 Db 229

Db 6006 ATGCAGCGCAATGCTCTACACCAACCGCGAACAAGCTGGTGGACGATATCACCGTGCTGA 5947
 QY 229 hrAspVal-----AspGlyAsnAspGlyLeuIleAsnTyrgluValGluValAlaAsnG 247
 Db 5946 CGCATGTCGCGCAAGCTGTAAACACGACGCTCTGTACTGGCAGGTG---GTGGCCAAAT- 5891
 QY 247 lnrThrThrGlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleValAlaLysA 267
 Db 5890 -----CGTATGTCACCGCTTGAACCTCGATATCGGATCAACAGCGTGTTCACACG 5839
 QY 267 lAspGlyAlaGlnGlyThrValThrIleProSerValIleTyrGlnProGlyAlaA 287
 Db 5838 GACAAAGCACTAGCGGCACTTTGCAAGTGAATCCGACCTTCGGCAACCGGGCTGAAG 5779
 QY 287 lATyrLeuTyrGlnLeuGlnValaLeuIleValGlySerSerGlyAspValValAspThrT 307
 Db 5778 GTTATCTTATGAACCTGCGCTC-----ACAGCCAAAGCCAGACAGCTGATATCT 5725
 QY 307 yrAsnLeuAlaThrGlyValArgThrValIleValValaGlySerGlnPheLeuIleAsnG 327
 Db 5724 ACCCGCTTCGCGTCCGCGCATCCGCTGAGTGGAGGCGCAACGTTCCGTATTAAC 5665
 QY 327 lYLeuProPheTyrPheThrGlyPheGlyLysHisGluAspThrAlaValArgGlyLysG 347
 Db 5664 ACAAAACCGTTCTTACTTACTGCTTGGTGTGTCATGAATCGGCACTTACGTGGCAAG 5605
 QY 347 lYHisAspProAlaTyrMetValHisAspPheGlnLeuMetLysTrpIleGlyAlaAsnS 367
 Db 5604 GATTGATTAACGTCCTGATGTGACAGACACGCACTTAATGACCTGGATTGGGCGCAACT 5545
 QY 367 ePheArgThrSerHisTyrProTyraAlaGluGlnValMetAspPheAlaAspArgAsnG 387
 Db 5544 CTAACGCTACCTCGCATTAACCTTAAGTGAAGAGATGCTGACCTGGCGGCAATGAACAG 5485
 QY 387 lYIleValIleIleAspGluThrProAlaValaGlyLeuAsnIleAlaLeu---MetGlyV 406
 Db 5484 GCATCGTGTGATGATGTAAGACTGCTGCTGCTTAACTCTCTTAAAGCATGTGTT 5425
 QY 406 alSerGluSerGlyAlaPro---GlnThrPheThrProAspAlaIleAsnAspLysThrG 425
 Db 5424 TCGAAGCGGCAACAAGCGGAAGAACTGTACAGCAAGCAAGCAAGCGGGAACATC 5365
 QY 425 lngIluAlaHisIleArgIleAlaIleArgGluLeuIleAlaIleArgAspLysAsnHisAlaSerV 445
 Db 5364 ACAAGCGCATTAACGCGGATTAAGAGCTATAGCGGCTGACAAACCAACCCAGCG 5305
 QY 445 alValMetTrpSerIleAlaAsnGluProAlaSerHisGluAspGlyAlaArgGlyTyrP 465
 Db 5304 TGTGATGTGGAGTATTTGCCAAGAACCGGATACCGCTCCGACAGGTGCACGGGAATAT 5245
 QY 465 heGluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaA 485
 Db 5244 TCGCGCCACTGGCGGAAGCAAGCGTAATGACCCGAGCGGTCCGATCACCTGCGCTGA 5185
 QY 485 snValGlyThrAlaThrTyrglnLeuAspArgIleSerAspLeuPheAspValIleSerCysI 505
 Db 5184 ATGTATATGTTCCGACGCTCACACCGAATACATACGATCTCTTTGATGTGCTGTGCC 5125
 QY 505 lAsnAspGlyPheGlyTyrTyrglnThrGlyAspLeuGluAlaGluAlaIle 525
 Db 5124 TGAACCGTTATACCGATGTGTATGCCAAGCGCGGATTTGAAACCGCGAGAAAGTATC 5065
 QY 525 euGlyLysGlyLeuHisGlyTyrGlnGlyLysPheHisAspProIleValMetThrGluT 545
 Db 5064 TGAATAAACAATCTTCGCGCTGACAGGAATACTGATCAGCGCATATATCATCAACCGAAT 5005
 QY 545 yrcGlyAlaAspThrIleuAlaGlyLeuHisSerIleLeuGlyLeuProTrpSerGluGluP 565
 Db 5004 ACGGGGTGATACGTTACCGCGGCTGCACTGAATGACACCGAATGATGAAAGACT 4945
 QY 565 heGlnValGlnMetLeuAspMetTyrgHisArgValPheAspArgIleGluSerMetAlaG 585
 Db 4944 ATCAGTGTGACAGTGTGATGTATCAACCGGCTTGTGATGCGGTCAAGCGCGCTGCTCG 4885

QY 585 |ygluHhIvAl|TrrpAnpheAlaaspPhcglIntrhAnleuglYllleArGValAspG 605
 DB 4884 GTGAACAGGATGGAATTCGCGGATTTGCGACCTCGGACGACATATTCGCGGTTGGCG 4825
 QY 605 |yAsnlyeLyvgGlYVal|PheThrArGAspArGlYsProlYsAlaAlaAlhISerLeuA 625
 DB 4824 GTTACAGAGAAAGGATCTTCACTCGGACCGCAACCGAAGTGTGGCGGCTTTCTGTCTGC 4765
 QY 625 rglAlaArGTrpThrSerlleArp 632
 DB 4764 AAAACGCTGACGATGATGAC 4742

RESULT 7
 AY237648/c 11973 bp DNA circular SYN 19-APR-2003
 LOCUS AY237648
 DEFINITION Cloning vector pHR50, complete sequence.
 ACCESSION AY237648
 VERSION AY237648.1 GI:30025585
 KEYWORDS
 SOURCE Cloning vector pHR50
 ORGANISM Cloning vector pHR50
 OTHER SEQUENCES: artificial sequences; vectors.
 REFERENCE 1 (bases 1 to 11973)
 AUTHORS Ramos,H.J.O., Soares-Ramos,J.R.L., Souza,E.M. and Pedrosa,F.O.
 TITLE Submitted (14-FEB-2003) Department of Biochemistry and Molecular
 BIOLOGY, Universidade Federal do Parana - UFPR, Centro Politecnico,
 Curitiba, Parana 81531-990, Brazil
 LOCATION/Qualifiers
 1. 11973
 /organism="Cloning vector pHR50"
 /mol_type="genomic DNA"
 /db_xref="taxon:227048"
 1. 2642
 /note="mob and rep regions from broad-host-range vector
 pBBR1-GFP deposited in GeneBank Accession number AF110459"
 complement(2882..3776)
 /gene="aph"
 /note="synonym: nptII; derived from pMM6"
 complement(2982..3776)
 /gene="aph"
 /note="NPTII kanamycin resistance gene"
 /codon_start=1
 /transl_table=11
 /product="aminoglycoside-3'-O-phosphotransferase"
 /protein_id="AAP04589.1"
 /db_xref="GI:30025556"
 /translation="MIEODGLHAGSPAAWVERLFGYDMAQQTIGSDAAVFRLSAQGR
 PVLFVKTDSGALNELQDEAARLSMTATTCGCAVLADVTATGRMDLILGVRQDL
 LSHSLAPAEKVSIMADMRRLHTLDTATCFPDQAKRIERATRRKAGLVDODDDE
 EHQGLAPAEI/PALKAKMPDGEDLVVTHGDACLPNTMVENGRPSGFIIDCGRLGVAURY
 ODIALATRDIAEELGEMADRFVLVGIAPDSQRIAFYRLDEFP"
 complement(4706..6517)
 /gene="uidA"
 /note="derived from pMM6; expressed from gentamycin
 resistance aacCI promoter"
 complement(4706..6517)
 /gene="uidA"
 /codon_start=1
 /transl_table=11
 /product="beta-glucuronidase"
 /protein_id="AAP04590.1"
 /db_xref="GI:30025557"
 /translation="MVRPVETPTPREIKLGLMAFSLDRENCIDQRMESALQESRA
 IAVGSPNDQFADADIRYAGNVMYOREVEIPKMGAGORIVLFDVATVYTGKVMQO
 EWNEHOGGYTPFADAVTPYIAGKSVITTCVNNELMOTIPGAVITDENGKKOXY
 FHDFFVAGIHSBVMVLTPTMTWDDITVTHAOCNHSADVMOVANGDVSELRD
 ADOQVAVTGCSTGLQVYVPHMKGEGEIVLCTTAQSQTCDIYPLRVGRSTAY
 KGGQPLINRPFYFTFGRHEDADLKGKGDVNLMDHDLMDWIGANSRTRISHIYA
 EEWLADNDEHGIIVIDETAVALNLSLIGFGFAGNPKPELYSEBAANGESTRQOHLAI
 KELLARDKHPVSVVMVSIANEPTRQGAAREYAPLAZATRKLDPTPRITCVMMFC
 AHTDTISDLFDVLCINRYGVMYQSGDLFETAKEVLEKELLAMQEKLHQPILITREYVD

misc_feature
 /note="mob and rep regions from broad-host-range vector
 pBBR1-GFP deposited in GeneBank Accession number AF110459"
 complement(2882..3776)
 /gene="aph"
 /note="synonym: nptII; derived from pMM6"
 complement(2982..3776)
 /gene="aph"
 /note="NPTII kanamycin resistance gene"
 /codon_start=1
 /transl_table=11
 /product="aminoglycoside-3'-O-phosphotransferase"
 /protein_id="AAP04589.1"
 /db_xref="GI:30025556"
 /translation="MIEODGLHAGSPAAWVERLFGYDMAQQTIGSDAAVFRLSAQGR
 PVLFVKTDSGALNELQDEAARLSMTATTCGCAVLADVTATGRMDLILGVRQDL
 LSHSLAPAEKVSIMADMRRLHTLDTATCFPDQAKRIERATRRKAGLVDODDDE
 EHQGLAPAEI/PALKAKMPDGEDLVVTHGDACLPNTMVENGRPSGFIIDCGRLGVAURY
 ODIALATRDIAEELGEMADRFVLVGIAPDSQRIAFYRLDEFP"
 complement(4706..6517)
 /gene="uidA"
 /note="derived from pMM6; expressed from gentamycin
 resistance aacCI promoter"
 complement(4706..6517)
 /gene="uidA"
 /codon_start=1
 /transl_table=11
 /product="beta-glucuronidase"
 /protein_id="AAP04590.1"
 /db_xref="GI:30025557"
 /translation="MVRPVETPTPREIKLGLMAFSLDRENCIDQRMESALQESRA
 IAVGSPNDQFADADIRYAGNVMYOREVEIPKMGAGORIVLFDVATVYTGKVMQO
 EWNEHOGGYTPFADAVTPYIAGKSVITTCVNNELMOTIPGAVITDENGKKOXY
 FHDFFVAGIHSBVMVLTPTMTWDDITVTHAOCNHSADVMOVANGDVSELRD
 ADOQVAVTGCSTGLQVYVPHMKGEGEIVLCTTAQSQTCDIYPLRVGRSTAY
 KGGQPLINRPFYFTFGRHEDADLKGKGDVNLMDHDLMDWIGANSRTRISHIYA
 EEWLADNDEHGIIVIDETAVALNLSLIGFGFAGNPKPELYSEBAANGESTRQOHLAI
 KELLARDKHPVSVVMVSIANEPTRQGAAREYAPLAZATRKLDPTPRITCVMMFC
 AHTDTISDLFDVLCINRYGVMYQSGDLFETAKEVLEKELLAMQEKLHQPILITREYVD

gene
 /note="mob and rep regions from broad-host-range vector
 pBBR1-GFP deposited in GeneBank Accession number AF110459"
 complement(2882..3776)
 /gene="aph"
 /note="synonym: nptII; derived from pMM6"
 complement(2982..3776)
 /gene="aph"
 /note="NPTII kanamycin resistance gene"
 /codon_start=1
 /transl_table=11
 /product="aminoglycoside-3'-O-phosphotransferase"
 /protein_id="AAP04589.1"
 /db_xref="GI:30025556"
 /translation="MIEODGLHAGSPAAWVERLFGYDMAQQTIGSDAAVFRLSAQGR
 PVLFVKTDSGALNELQDEAARLSMTATTCGCAVLADVTATGRMDLILGVRQDL
 LSHSLAPAEKVSIMADMRRLHTLDTATCFPDQAKRIERATRRKAGLVDODDDE
 EHQGLAPAEI/PALKAKMPDGEDLVVTHGDACLPNTMVENGRPSGFIIDCGRLGVAURY
 ODIALATRDIAEELGEMADRFVLVGIAPDSQRIAFYRLDEFP"
 complement(4706..6517)
 /gene="uidA"
 /note="derived from pMM6; expressed from gentamycin
 resistance aacCI promoter"
 complement(4706..6517)
 /gene="uidA"
 /codon_start=1
 /transl_table=11
 /product="beta-glucuronidase"
 /protein_id="AAP04590.1"
 /db_xref="GI:30025557"
 /translation="MVRPVETPTPREIKLGLMAFSLDRENCIDQRMESALQESRA
 IAVGSPNDQFADADIRYAGNVMYOREVEIPKMGAGORIVLFDVATVYTGKVMQO
 EWNEHOGGYTPFADAVTPYIAGKSVITTCVNNELMOTIPGAVITDENGKKOXY
 FHDFFVAGIHSBVMVLTPTMTWDDITVTHAOCNHSADVMOVANGDVSELRD
 ADOQVAVTGCSTGLQVYVPHMKGEGEIVLCTTAQSQTCDIYPLRVGRSTAY
 KGGQPLINRPFYFTFGRHEDADLKGKGDVNLMDHDLMDWIGANSRTRISHIYA
 EEWLADNDEHGIIVIDETAVALNLSLIGFGFAGNPKPELYSEBAANGESTRQOHLAI
 KELLARDKHPVSVVMVSIANEPTRQGAAREYAPLAZATRKLDPTPRITCVMMFC
 AHTDTISDLFDVLCINRYGVMYQSGDLFETAKEVLEKELLAMQEKLHQPILITREYVD

CDS
 /note="mob and rep regions from broad-host-range vector
 pBBR1-GFP deposited in GeneBank Accession number AF110459"
 complement(2882..3776)
 /gene="aph"
 /note="synonym: nptII; derived from pMM6"
 complement(2982..3776)
 /gene="aph"
 /note="NPTII kanamycin resistance gene"
 /codon_start=1
 /transl_table=11
 /product="aminoglycoside-3'-O-phosphotransferase"
 /protein_id="AAP04589.1"
 /db_xref="GI:30025556"
 /translation="MIEODGLHAGSPAAWVERLFGYDMAQQTIGSDAAVFRLSAQGR
 PVLFVKTDSGALNELQDEAARLSMTATTCGCAVLADVTATGRMDLILGVRQDL
 LSHSLAPAEKVSIMADMRRLHTLDTATCFPDQAKRIERATRRKAGLVDODDDE
 EHQGLAPAEI/PALKAKMPDGEDLVVTHGDACLPNTMVENGRPSGFIIDCGRLGVAURY
 ODIALATRDIAEELGEMADRFVLVGIAPDSQRIAFYRLDEFP"
 complement(4706..6517)
 /gene="uidA"
 /note="derived from pMM6; expressed from gentamycin
 resistance aacCI promoter"
 complement(4706..6517)
 /gene="uidA"
 /codon_start=1
 /transl_table=11
 /product="beta-glucuronidase"
 /protein_id="AAP04590.1"
 /db_xref="GI:30025557"
 /translation="MVRPVETPTPREIKLGLMAFSLDRENCIDQRMESALQESRA
 IAVGSPNDQFADADIRYAGNVMYOREVEIPKMGAGORIVLFDVATVYTGKVMQO
 EWNEHOGGYTPFADAVTPYIAGKSVITTCVNNELMOTIPGAVITDENGKKOXY
 FHDFFVAGIHSBVMVLTPTMTWDDITVTHAOCNHSADVMOVANGDVSELRD
 ADOQVAVTGCSTGLQVYVPHMKGEGEIVLCTTAQSQTCDIYPLRVGRSTAY
 KGGQPLINRPFYFTFGRHEDADLKGKGDVNLMDHDLMDWIGANSRTRISHIYA
 EEWLADNDEHGIIVIDETAVALNLSLIGFGFAGNPKPELYSEBAANGESTRQOHLAI
 KELLARDKHPVSVVMVSIANEPTRQGAAREYAPLAZATRKLDPTPRITCVMMFC
 AHTDTISDLFDVLCINRYGVMYQSGDLFETAKEVLEKELLAMQEKLHQPILITREYVD

gene
 /note="mob and rep regions from broad-host-range vector
 pBBR1-GFP; expressed from gentamycin
 resistance aacCI promoter"
 complement(6610..7326)
 /gene="gfp"
 /note="derived from pBBR1-GFP; expressed from gentamycin
 resistance aacCI promoter"
 complement(6610..7326)
 /gene="gfp"
 /note="modified form multi; derived from Agrobacterium victoria
 GFP deposited in GeneBank Accession number M62653"
 /codon_start=1
 /transl_table=11
 /product="green fluorescent protein"
 /protein_id="AAP04593.1"
 /db_xref="GI:30025560"
 /translation="MSKGEELFTGCVPLIVELDGVNKGKPSVSGREGDATTGKTL
 KRITCTTGKLPVWPFLVTFYGVQVCFAPYPMHMQHDFKAMPEGVQERTIPKD
 DSNYKTRAEKFECDTLVNRILKIDFEDENILGHLKLEYNSHNVIMADKONG
 IKVNFKIRNIEDSGYQLADHYQONTPIGDGVLLPDNNHLYSTQSALSKDPNEKRDHM
 VLEFVTAAGITHGMDLYK"
 complement(7376..8950)
 /gene="nifA"
 /note="derived from Klebsiella pneumoniae nifA gene of
 pNH1; expressed from gentamycin resistance aacCI
 promoter"
 complement(7376..8950)
 /gene="nifA"
 /codon_start=1
 /transl_table=11
 /product="nif gene transcriptional activator"
 /protein_id="AAP04591.1"
 /db_xref="GI:30025558"
 /translation="MIRKSDSDPTTARRPDLSSQFTAMORISVILSPATASKTIOBLV
 SYLNDARFMOHMCILYDSQOEILSIEBALQCEBDQTLRSTQIRPRPGSGVITVLAQ
 GQSLVLRVADQDFLDRSLYDYDLPFLAVLMGHSRPPGLVLAHAARDEBRIPA
 CRFLFETVNLIAQITRLMILPTSAQAQOSPRIERPACPTSRFGLENNVQKSPA
 MROIMDIRQVSRMDTVLVRSQSGKELINAIHNSPPRAAFAVKENCALPNTL
 LSELFGHEKGAFTGAVROKRFELADGTLFLDITGSSASFOKALRIQEGEME
 RVGDEPTLAVNRIIAATRRHLEBRVLCHPEDYFVRLNMPIALPLRQEDQIAE
 LMFVLRKIAHQGRTLRISDQATIRLMKYSFGNVRLENCIERAVLSSEGLIRD
 VLFNHRDNPRLASSGPAEDGMDLNSLDERORLIALLEKAGWQAKARLLGMTPR
 QVARIQIMDIMTPRL"
 8951..10298
 /note="derived from pQ200Ks"
 complement(9520..9525)
 /note="derived from gentamycin resistance aacCI gene of
 pQ200Ks"
 complement(9548..9553)
 /note="derived from gentamycin resistance aacCI gene of
 pQ200Ks"
 complement(10902..11762)
 /gene="dia"
 /note="derived from pBBR1-GFP"
 complement(10902..11762)
 /gene="dia"
 /codon_start=1
 /transl_table=11
 /product="beta-lactamase"
 /protein_id="AAP04592.1"
 /db_xref="GI:30025559"
 /translation="MSIOHFRVALIPFFAFCLPVPAREPTELVKVADQDLGARVGY
 IRLDNGSKILSPREPERPMMSTGFVLCGAVIRIDAGQOLGRRILHYSQNDLVE
 YSPVTRKHLTDKMTARELSAALITMSDNTANALLTTTGKPRELTAFLNMDGVYRL
 DRWEPLEAIENFDERDITWVAMATTKLTLGELTLASROQLIDNMEADVYACPL
 LNSALPWGFIIDKSGAGRGSGIILAGPDKSRIVIVITTSQATWDERNROIA
 EIGASLIKHM"

misc_feature
 /note="mob and rep regions from broad-host-range vector
 pBBR1-GFP; expressed from gentamycin
 resistance aacCI promoter"
 complement(6610..7326)
 /gene="gfp"
 /note="derived from pBBR1-GFP; expressed from gentamycin
 resistance aacCI promoter"
 complement(6610..7326)
 /gene="gfp"
 /note="modified form multi; derived from Agrobacterium victoria
 GFP deposited in GeneBank Accession number M62653"
 /codon_start=1
 /transl_table=11
 /product="green fluorescent protein"
 /protein_id="AAP04593.1"
 /db_xref="GI:30025560"
 /translation="MSKGEELFTGCVPLIVELDGVNKGKPSVSGREGDATTGKTL
 KRITCTTGKLPVWPFLVTFYGVQVCFAPYPMHMQHDFKAMPEGVQERTIPKD
 DSNYKTRAEKFECDTLVNRILKIDFEDENILGHLKLEYNSHNVIMADKONG
 IKVNFKIRNIEDSGYQLADHYQONTPIGDGVLLPDNNHLYSTQSALSKDPNEKRDHM
 VLEFVTAAGITHGMDLYK"
 complement(7376..8950)
 /gene="nifA"
 /note="derived from Klebsiella pneumoniae nifA gene of
 pNH1; expressed from gentamycin resistance aacCI
 promoter"
 complement(7376..8950)
 /gene="nifA"
 /codon_start=1
 /transl_table=11
 /product="nif gene transcriptional activator"
 /protein_id="AAP04591.1"
 /db_xref="GI:30025558"
 /translation="MIRKSDSDPTTARRPDLSSQFTAMORISVILSPATASKTIOBLV
 SYLNDARFMOHMCILYDSQOEILSIEBALQCEBDQTLRSTQIRPRPGSGVITVLAQ
 GQSLVLRVADQDFLDRSLYDYDLPFLAVLMGHSRPPGLVLAHAARDEBRIPA
 CRFLFETVNLIAQITRLMILPTSAQAQOSPRIERPACPTSRFGLENNVQKSPA
 MROIMDIRQVSRMDTVLVRSQSGKELINAIHNSPPRAAFAVKENCALPNTL
 LSELFGHEKGAFTGAVROKRFELADGTLFLDITGSSASFOKALRIQEGEME
 RVGDEPTLAVNRIIAATRRHLEBRVLCHPEDYFVRLNMPIALPLRQEDQIAE
 LMFVLRKIAHQGRTLRISDQATIRLMKYSFGNVRLENCIERAVLSSEGLIRD
 VLFNHRDNPRLASSGPAEDGMDLNSLDERORLIALLEKAGWQAKARLLGMTPR
 QVARIQIMDIMTPRL"
 8951..10298
 /note="derived from pQ200Ks"
 complement(9520..9525)
 /note="derived from gentamycin resistance aacCI gene of
 pQ200Ks"
 complement(9548..9553)
 /note="derived from gentamycin resistance aacCI gene of
 pQ200Ks"
 complement(10902..11762)
 /gene="dia"
 /note="derived from pBBR1-GFP"
 complement(10902..11762)
 /gene="dia"
 /codon_start=1
 /transl_table=11
 /product="beta-lactamase"
 /protein_id="AAP04592.1"
 /db_xref="GI:30025559"
 /translation="MSIOHFRVALIPFFAFCLPVPAREPTELVKVADQDLGARVGY
 IRLDNGSKILSPREPERPMMSTGFVLCGAVIRIDAGQOLGRRILHYSQNDLVE
 YSPVTRKHLTDKMTARELSAALITMSDNTANALLTTTGKPRELTAFLNMDGVYRL
 DRWEPLEAIENFDERDITWVAMATTKLTLGELTLASROQLIDNMEADVYACPL
 LNSALPWGFIIDKSGAGRGSGIILAGPDKSRIVIVITTSQATWDERNROIA
 EIGASLIKHM"

gene
 /note="mob and rep regions from broad-host-range vector
 pBBR1-GFP; expressed from gentamycin
 resistance aacCI promoter"
 complement(6610..7326)
 /gene="gfp"
 /note="derived from pBBR1-GFP; expressed from gentamycin
 resistance aacCI promoter"
 complement(6610..7326)
 /gene="gfp"
 /note="modified form multi; derived from Agrobacterium victoria
 GFP deposited in GeneBank Accession number M62653"
 /codon_start=1
 /transl_table=11
 /product="green fluorescent protein"
 /protein_id="AAP04593.1"
 /db_xref="GI:30025560"
 /translation="MSKGEELFTGCVPLIVELDGVNKGKPSVSGREGDATTGKTL
 KRITCTTGKLPVWPFLVTFYGVQVCFAPYPMHMQHDFKAMPEGVQERTIPKD
 DSNYKTRAEKFECDTLVNRILKIDFEDENILGHLKLEYNSHNVIMADKONG
 IKVNFKIRNIEDSGYQLADHYQONTPIGDGVLLPDNNHLYSTQSALSKDPNEKRDHM
 VLEFVTAAGITHGMDLYK"
 complement(7376..8950)
 /gene="nifA"
 /note="derived from Klebsiella pneumoniae nifA gene of
 pNH1; expressed from gentamycin resistance aacCI
 promoter"
 complement(7376..8950)
 /gene="nifA"
 /codon_start=1
 /transl_table=11
 /product="nif gene transcriptional activator"
 /protein_id="AAP04591.1"
 /db_xref="GI:30025558"
 /translation="MIRKSDSDPTTARRPDLSSQFTAMORISVILSPATASKTIOBLV
 SYLNDARFMOHMCILYDSQOEILSIEBALQCEBDQTLRSTQIRPRPGSGVITVLAQ
 GQSLVLRVADQDFLDRSLYDYDLPFLAVLMGHSRPPGLVLAHAARDEBRIPA
 CRFLFETVNLIAQITRLMILPTSAQAQOSPRIERPACPTSRFGLENNVQKSPA
 MROIMDIRQVSRMDTVLVRSQSGKELINAIHNSPPRAAFAVKENCALPNTL
 LSELFGHEKGAFTGAVROKRFELADGTLFLDITGSSASFOKALRIQEGEME
 RVGDEPTLAVNRIIAATRRHLEBRVLCHPEDYFVRLNMPIALPLRQEDQIAE
 LMFVLRKIAHQGRTLRISDQATIRLMKYSFGNVRLENCIERAVLSSEGLIRD
 VLFNHRDNPRLASSGPAEDGMDLNSLDERORLIALLEKAGWQAKARLLGMTPR
 QVARIQIMDIMTPRL"
 8951..10298
 /note="derived from pQ200Ks"
 complement(9520..9525)
 /note="derived from gentamycin resistance aacCI gene of
 pQ200Ks"
 complement(9548..9553)
 /note="derived from gentamycin resistance aacCI gene of
 pQ200Ks"
 complement(10902..11762)
 /gene="dia"
 /note="derived from pBBR1-GFP"
 complement(10902..11762)
 /gene="dia"
 /codon_start=1
 /transl_table=11
 /product="beta-lactamase"
 /protein_id="AAP04592.1"
 /db_xref="GI:30025559"
 /translation="MSIOHFRVALIPFFAFCLPVPAREPTELVKVADQDLGARVGY
 IRLDNGSKILSPREPERPMMSTGFVLCGAVIRIDAGQOLGRRILHYSQNDLVE
 YSPVTRKHLTDKMTARELSAALITMSDNTANALLTTTGKPRELTAFLNMDGVYRL
 DRWEPLEAIENFDERDITWVAMATTKLTLGELTLASROQLIDNMEADVYACPL
 LNSALPWGFIIDKSGAGRGSGIILAGPDKSRIVIVITTSQATWDERNROIA
 EIGASLIKHM"

ORIGIN
 Alignment Scores:
 Pred. No.: 2,056-115
 Score: 1718.50
 Percent Similarity: 69.004
 Best Local Similarity: 54.054
 Length: 11973
 Matches: 340
 Conservative: 94
 Mismatches: 172

Query Match: 51.24% Indels: 23
 DB: 12 Gaps: 10
 US-10-757-093-4 (1-634) x AY237648 (1-11973)

QY 18 GYTHPRQALALARGHIEPRQARGANGU--MetThrglnhisgluInpRoL 37
 DB 6597 GGAATTCCTGACGCCGGGGGATCCACAGAAACGTATGATGATTAAGCGCAAGCT 6538

QY 37 eu-----IlelySValargProglInargThrsSerargluVala 52
 DB 6537 TGGCTGACGTCAGTCCCTTATGTCGCTCTGTAGAAACCCCAACCGGTAAATCAAAA 6478

QY 52 snleuaspGlyLeuTrpLysPhealaLeu-----AlaSerGlyLeuasnaphra 69
 DB 6477 AACTGACGCGCTGGGCAATTCAGTCGATCCGCAAAACTGTGGATATGATCAG---- 6422

QY 69 laGlnProTrpThralaProleuProlysglyLeuGluCySProValProAlaSerTYR 89
 DB 6421 --CGTTGGTGGAAAGCCGCTTACAGAAAGCCGGCAATTGCTGTCAGGCGAGTTTAA 6364

QY 89 snApllePheIleSerargLuiIleHisaspHisValGlyTrpValTYrTrglnarg 109
 DB 6363 ACGATCACTTGGCCGATGACATATTCGTAATTAAGCGGCAACGTGTGATCAGCGCG 6304

QY 109 luValIleValProlysglyTrpSerGlnIuargTYrLeuValargAlaGluSerAlaT 129
 DB 6303 AAGCTTATATCCGAAAGGTGGGACAGCCGATGCTGCTGCTTCGATTCGGCTCA 6244

QY 129 hrihshiagIYargIleTYrValaLsnasnargLeuValAlaGluHisValGlyTYrT 149
 DB 6243 CTCATTACGGGAAAGTGGGTCAATATATCAGAAAGATGATGAGCATCAGGGGGCTATA 6184

QY 149 hrProPhegluaLaaspValTrhgIuLeuValAlaProglYgluLysPheargLeuThri 169
 DB 6183 CGCCATTGGAAGCGGATGATCAGCCGCGATGTATTCGCCGGGAAAGCTACGATCACCG 6124

QY 169 legIyValasnargluLeuThrihsgluThriIleProProglYlyIleThThrgIYA 189
 DB 6123 TTGTGTGAACAGAACTGAACCTGACGACTATCCCGCGGAAATGTGATTAAC--G 6067

QY 189 snAlaThrgIyLysargIleGlnThThrgIlnhisaspPheTYrAsnTYrAlaGlyLeu 209
 DB 6066 ACGAAACCGGCAAGAAAGACGCTTACTTCCATGATTTCTTAACTAGCGGATCC 6007

QY 209 laArgSerIleTrpLeuTYrSerValProglInhisIleGlnaspIleThValValT 229
 DB 6006 ATCGACGCTAATGCTCTACACCGCGCAACCTGGGTGACATATCACCGTGTGA 5947

QY 229 hrAspVal-----AspGlyAspAsnGlyLeuIleAsnTYrGluValAlaAsnG 247
 DB 5946 CGCATGTCGGGCAAGACTGTACACGCGCTGTGTGACTGGCAGGTG--GTGGCCAAAT- 5891

QY 247 lnThThrgIyGlnIleGlnIleSerValIleAspGluaspGlyAlaIleValAlaLys 267
 DB 5890 -----GGTATGTCAGCGGTGAACGTGCTATCCGATCAACAGGTTTGGCACTG 5839

QY 267 laSerGlyalagIyThValThriIleProSerValLysIleuTrpGlnProglYalaa 287
 DB 5838 GACAAAGGACTAGCGGCACTTTCAAGTGTGAAATCCGACCTGTGCAACCGGGGAAG 5779

QY 287 laTYrLeuTYrGlnLeuGlnValaAsnIleValGlySerSerGlyAspValValaspThT 307
 DB 5778 GTTATCTTATGAACTGTGCTC-----ACAGCCAAAGCCAGACAGATGTGATATCT 5725

QY 307 yzAsnleuAlaThrgIyValargThValIyValaIagIySerGlnPheleuIleAsnG 327
 DB 5724 ACCCGCTTGGCGGCGATCCGCTCAGTCACTGAGTGAAGGCCAAACGTTCTGATTAAC 5665

QY 327 lYlySProPheTYrPheThrgIyPheGlyLysHisgluAspThralaValargIyLysG 347
 DB 5664 ACAAACCGTTACTTACTGCTTGTGTGTCATGAATGCGGAGCTTAAGTGGCAAAAG 5605

QY 347 lYhIsapProAlaTYrMetValHisaspPheGlnLeuMetLysTrpIleGlyAlaAsn 367
 DB 5604 GATTTCGATAACGTCGATGCTGACAGCACCGCATTAATGACTGATGGGGCAACT 5545

QY 367 erPheargThSerhIstYrProTYrAlaGluGluValMetaspPheAlaaspArgAsnG 387
 DB 5544 CTACACGTCACCTGACATTAACCTCTGACGAAAGATGCTCGACATGGGCAATGAATG 5485

QY 387 lYlIleValIleAspGluThrProAlaValaGlyLeuasnIlealaLeu--MetGlyV 406
 DB 5484 GCATCGGTGATGATGATAAAGTGTGCTGTGGCTTTAAGCTCTCTTAAGGATGTGTT 5425

QY 406 alSerGluSerGlyAlaPro--GlnThPheThrProaspAlaIleasnApLyThG 425
 DB 5424 TCAGACGGGCAACAAGCCGAAAGAACTGTGACGCAAGAGGACAGTCAACGGGAAATC 5365

QY 425 lngluAlahislysglnAlaIleargGluLeuIleAlaargAspLysAsnHisAlaSerY 445
 DB 5364 AGCAACGCGACTTACAGCGGATTAAGAGCTGATAGCGGTGACAAACCAACCAAGCG 5305

QY 445 alValMetTrpSerIleAlaasnGluProAlaSerHisgluAspGlyAlaargIuTYrP 465
 DB 5304 TGGTATGTGGATATGTCGCAACGAAACCGATATCCGTCGCAAGTGCACGGGATATT 5245

QY 465 heGluProleuThrasnleuThrargGlnLeuaspProThrargProIleThrPheAla 485
 DB 5244 TGCCGCACTGGCGGAGCAAGCGGTAACCTGACCCGACGCGTCCGATCACCTGCTCA 5185

QY 485 snValGlyThralaThrTYrGlnLeuAspArgIleSeraspLeuPheaspValSerCyS 505
 DB 5184 ATGTATATGTTTCGCGACGCTCACACCGATACATCAGGATCTCTTATGTGCTGTCC 5125

QY 505 laasnargTYrPheGlyTrpTYrSerGlnThrgIyAspLeuGluValaGluAlaIal 525
 DB 5124 TGAACGTTATTAAGATGTATGTCGCAAGCGGATTTGGAAACGCAAGAAAGTAC 5065

QY 525 euGluLysgluLeuLeuIleGlyTrpGlnGluLysPheHisargProIleValMetThrgI 545
 DB 5064 TGGAAAAAGAACTTGTGGCGCTGCGAGAAAGTGCATCAGCCGATTTATCACCGAAT 5005

QY 545 yrgIyAlaaspThrleuAlaGlyLeuHisSerIleleuGlyLeuProTrpSerGluIup 565
 DB 5004 ACGGGGTGATACGTTTACCGGGCTGCACCTCAATGTACAGCACATGGAGTGAAGGT 4945

QY 565 heGlnValaGlnMetLeuaspMetTYrHisargValaPheaspArgIleGluSerMetAla 585
 DB 4944 ATCATGTGCATGCTGATATGATATCCGCGCTTGTGATGCGCGCGCTGTG 4885

QY 585 lYgluHisValTrpAsnPheAlaaspPheGlnThrasnleuGlyIleIleargValaasp 605
 DB 4884 GTGAACAGGTATGGAATTTCCCGCATTTTGGCACTGCGCAAGGCAATATGCGCTTGGCG 4825

QY 605 lYAsnIyLysGlyValaPheThrargAspArgIyProlyValaAlaAlaHisSerLeu 625
 DB 4824 GTAAACAAAGGATCTTCACTGCGAACCGAAAGTCCGCGGCTTTCTGCTGC 4765

QY 625 rgaLaargTrpThSerIleasp 632
 DB 4764 AAAAAGCTGACCTGGCATGAAC 4742

RESULT 8
 AF305918 2103 bp DNA linear BCT 19-SEP-2001
 LOCUS Escherichia coli O157:H- beta-glucuronidase (uidA) gene, complete
 DEFINITION cde
 ACCESSION AF305918
 VERSION AF305918.1 GI:13194573
 KEYWORDS
 SOURCE
 ORGANISM
 Escherichia coli
 Escherichia coli
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.
 REFERENCE
 1 (bases 1 to 2103)

AUTHORS Monday, S.R., Whitlam, T.S. and Feng, P.C.
TITLE Genetic and evolutionary analysis of mutations in the gusA gene that cause the absence of beta-glucuronidase activity in *Escherichia coli* O157:H7
JOURNAL J. Infect. Dis. 184 (7), 918-921 (2001)
MEDLINE 21434142
PUBMED 11510000
REFERENCE 2 (bases 1 to 2103)
AUTHORS Monday, S.R. and Feng, P.
TITLE Direct Substitution
JOURNAL Submitted (15-SEP-2000) U.S. Food and Drug Administration, 200 C St. SW, Washington, DC 20204, USA
FEATURES Location/Qualifiers
 source
 1. 2103
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /strain="493b1"
 /serotype="O157:H-"
 /db_xref="taxon:562"
 /country="Germany"
 /note="sorbitol fermenting strain"
 179. 1990
 /gene="uidA"
 /note="synonym: gusA"
 179. 1990
 /codon_start=1
 /transl_table=11
 /product="beta-glucuronidase"
 /protein_id="AAK15468.1"
 /db_xref="GI:13194574"
 /translation="MLRPVETPTREIKKLDGLAFSLDRENCIEORWMSALQESRA
 IVPGSFNDOPADDIRNVGVWVQREVFIPKMGAGQRTIVRFDLVTHYGVWVNO
 EVMHOGCTPREADVTTPYVIAKSKSRITVCNNELNQTTPGAVITTBENKSKOST
 FHDFFVAGIHSVLMVLYTPNTWDDIVVTHVADCNASVDMOVANQVDSVELRD
 ADQOVATQGTSGTLQVVA PHLMQEGYLYELVYTAQSQCEDLYPRVGRSHYVA
 KEOQFLINKPPTFTGFRHEDADLKGFDVLMVHDLMDMIGANSYRSHYVA
 BELMDADHGIIVIDETAAVGNLSLIGFEGANKPKELYSEAVANGTOOHLQAI
 KELIARDKHPSPVMSIANEPDTPROGAREVPLAEATRLDPRPTTCNVWMD
 AHTPTISDLPVLCINRYGKVVQSGDLTAEVLEKELIANQEKHQIITTBVVD
 TLAGHSKTDMSSEYOCAMDMTHRVFDRVSAVVEGVNMFADPATISQGLRVGN
 KGIIFRDKPKSAFLQKRWTKGNFGEKPOGGKQ"

ORIGIN

Alignment Scores:

Pred. No.:	2,33e-116	Length:	2103
Score:	1718.00	Matches:	333
Percent Similarity:	69.594	Conservative:	95
Best Local Similarity:	54.154	Mismatches:	169
Query Match:	51.224	Indels:	18
DB:	1	Gaps:	9

US-10-757-093-4 (1-634) x AF305918 (1-2103)

23 ArgHisPheProArgSerGluMetThrGlnHisGluGlnProLeuLeuIleValArgPro 42
 149 CAGTATTTTATCTTAAATGAGAGTC-----CTTATGTTACTCTCT 190

43 GlnArgThrSerSerArgGluLeuValAenLeuAspGlyLeuTrpLysPheAlaLeu--- 61
 191 GTAGAAACCCCAACCGGTGAATCAAAAATCTGACGGCCTGTGGCATTGAGTCTGAT 250

62 AlaSerGlyLeuAsnAspThrAlaGlnProTrpThrAlaProLeuProIleGlyLeuGlu 81
 251 CGCGAAACTGTGGAAATTGACGACCTGTGGGAAAGCCGTTACAAAGAACCCGGCA 310

82 CysProValProAlaSerTyrAsnAspIlePheIleSerArgGluIleHisAspHisVal 101
 311 ATTGCTGTCCAGCGAGCTTTAAAGATTCAGCTTCCGCGATGACATATTTGTAATTATG 370

102 GlyTrpValTyrTyrGlnArgGluValIleValProIleGlyTyrSerGlnGluArgTyr 121
 371 GGCAACGTCTGGTATCAGCGCAAGTCTTATACCAAGATTGGGACGACGCTATC 430

122 LeuValArgAlaGluSerAlaThrHisGlyValArgIleTyrValAsnAspArgLeuVal 141
 431 GTCTCGCTTTCATCGGCTCACTTACGCAAAAGTGGGTCAATATATCGAAGT 490

142 AlaGluHisValGlyTyrThrProPheGluAlaAspValThrGluLeuValAlaPro 161
 491 ATGACATCAGGAGCGGCTATACGATTTGAAGCCGATGTCAAGCGGTATATGCC 550

162 GlyGluLysPheArgLeuThrIleGlyValAsnAspGluLeuThrHisGluThrIlePro 181
 551 GGGAAAAGTGTACAGTATCAGTTGTGACACAAAGTAACTGAATCGGACGATATCCG 610

182 ProGlyLysIleThrThrArgLysAsnAlaThrGlyLysValIleGlnThrGlnHisAsp 201
 611 CGGGAAATGTGATTTACC---GACGAAACGGCAAGAAAAGCATTTTCTTCATGAT 667

202 PheTyrAsnTyrAlaGlyLeuAlaArgSerIleTrpLeuTyrSerValProGlnGlnHis 221
 668 TTCTTTATCAGCGCGGATCCATCGACGCTAATCTGTACACCAAGCGCAACCTGG 727

222 IleGlnAspIleThrValValThrAspVal-----AspGlyAspAsnGlyLeuIleAsn 239
 728 GTGGACGATATCACCTGTGTGACGATGTGCGCAAGACTGTAAACCAAGCGTCTGTGAC 787

240 TyrGluValGluValAlaAsnGlnThrThrGlnIleGlnIleSerValIleAspGlu 259
 788 TGCGAGGTG---GTGGCAAT-----GGATGTACAGCTTAACTGCGTGAATGCC 835

260 AspGlyAlaIleValAlaLysAlaSerGlyAlaGlnGlyThrValThrIleProSerVal 279
 836 GATCAACAGGTGTTGCACTGGACAAAGGACACAGCGGACTTGCAGAGCGTGAATCCG 895

280 LysLeuTrpGlnProGlyAlaAlaTyrLeuTyrGlnLeuGlnValAsnIleValGlySer 299
 896 CACCTGTGCAACCGGCTAAGGTATCTCTATGAACCTGACTC-----ACAGCCA 949

300 SerGlyAspValValAspThrTyrAsnLeuAlaThrGlyValArgThrValLysValAla 319
 950 AGCCACAGAGTGTATATCTACCCGTCGCGGCTGCGGATCCGGTCAGGCGAGTGAAG 1009

320 GlySerGlnPheLeuIleAsnGlyLysProPheTyrPheThrGlyPheGlyLysIleGlu 339
 1010 GGGCAACACTTCTCTATCAACCAACACCGTTCTACTGCGCTTGGCGCTATGA 1069

340 AspThrAlaValArgGlyLysGlyHisAspProAlaTyrMetValHisAspPheGlnLeu 359
 1070 GATGCGGATTTCCGCGCAAGGATTCGATACGTCGATGTCACGATCAGCATTA 1129

360 MetLysTrpIleGlyValAsnSerPheArgThrSerHisTyrProTyrAlaGluGluVal 379
 1130 ATGAGTGGATTTGGGCAACTCTACCGTACCTGCACTTACCTTACCTGAAGAGAT 1189

380 MetAspPheAlaAspArgAsnGlyIleValIleAspGlyThrProAlaValGlyLeu 399
 1190 CTGCACTGGCGAGATCAATGATCGATCGATGATGAAGAACTCCACTGCGCTTT 1249

400 AsnIleAlaLeu---MetGlyValSerGluSerGlyAlaPro---GlnThrPheThrPro 417
 1250 AACCTCTCTTATGAGCATTTGTTTCGAAAGCGGCAACCAACCAAGAACTGTACAGCGAA 1309

418 AspAlaIleAsnAspLysThrGlnGluAlaHisLysGlnAlaIleArgGluLeuIleAla 437
 1310 GAGGCAAGTCAACGGGGAACCTACGAGCGCACTTACAGCGCATTTAAAGAGCTGATAGCG 1369

438 ArgAspLysAsnHisAlaSerValValMetTyrSerIleAlaAsnGluProAlaSerHis 457
 1370 CGTGACAAACCAACCAAGCTGTGTGATGTGATGATGATCCAAAGAACCGATACCCCT 1429

458 GluAspGlyAlaArgGluTyrPheGluProLeuThrAsnLeuThrArgGlnLeuAspPro 477
 1430 CCGCAAGGTGCAAGGGAATATTTCCGCACTGCGGGAAGCAACCGTAACTCGATCCG 1489

QY 478 ThrArgProIleThrPheAlaIasnValGlyThrAlaThrTyrGlnLeuAspArgIleSer 497
 DB 1490 AGCGCGCCATCATCCGCTCAATGATGTTCTGGACGCTCACCCATCCATCCATCAC 1549
 QY 498 AspLeuPheAspValSerCysIleAenArgTyrPheGlyTyrPheSerGlnThrGlyAsp 517
 DB 1550 GATCTCTTGAATGTGTGTGCTGCAACCTTATACGTTGGTATGTCGCAAGCGCGCAT 1609
 QY 518 LeuGluGluAlaGluAlaLeuGluGlySerGluLeuHisGlyTyrGlnGluLysPheHis 537
 DB 1610 TTGGAAACCGCAGAGAAAGTACTGGAAAAAGAACTTCTGGCTGGCAGAGAAAGCTGCAT 1669
 QY 538 ArgProIleValMetThrGlnTyrGlyAlaAspThrLeuAlaGlyLeuHisSerIleLeu 557
 DB 1670 CAGCGCATTAATCATCCCAATACCGCGTGATACCTTAGCCGGCTGCATCAATGATAC 1729
 QY 558 GlyLeuProTyrSerGluGluPheGlnValGlnMetLeuAspMetTyrHisArgValPhe 577
 DB 1730 ACCGACATGTGAGTGAAGATATCAGTGTGATGCTGCTGATGTATCACCAGCGCTTT 1789
 QY 578 AspArgIleGluSerMetAlaGlyGluHisValTyrAsnPheAlaAspPheGlnThrAsn 597
 DB 1790 GATCGCGTACGCGCCCTGCTGCTGAACAGATATGAAATTCGCGCATTTTGGCATCTGC 1849
 QY 598 LeuGlyIleIleArgValAspGlyAsnLysGlyValPheThrArgAspArgLysPro 617
 DB 1850 CAAGGCATATGCGCGCTTGGCGGTACAGAAAGGAGATCTTACCCGCGCAACCGCAACCG 1909
 QY 618 LysAlaAlaAlaHisSerLeuArgAlaArgTyrPheSerIleAsp 632
 DB 1910 AAGTCGCGCGCTTTCTGCTGCMAAAACGCTGACATGCATGAC 1954
 RESULT 9
 AX025617 4299 bp DNA linear PAT 16-SEP-2000
 LOCUS AX025617
 DEFINITION Sequence 17 from Patent WO0029592.
 ACCESSION AX025617
 VERSION AX025617.1 GI:10187285
 KEYWORDS
 SOURCE
 ORGANISM
 1
 Logemann, E., Somsich, I., Hahlbrock, K., Kirsch, C. and Rushton, P.
 Chimeric promoters capable of mediating gene expression in plants
 upon pathogen infection and uses thereof
 Patent: WO 0029592-A 17 25-MAY-2000;
 MAX PLANCK GEBILDSCHAFT (DE) ; LOGEMANN ELKE (DE) ; SOMSICH IMRE
 (DE) ; HAHNBROCK KLAUS (DE) ; KIRSCH CHRISTOPH (DE) ; RUSHTON PAUL
 (GB)
 FEATURES
 source location/Qualifiers
 1..4299
 /organism="synthetic construct"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32630"
 /note="synthetic, no natural origin"
 ORIGIN
 Alignment Scores:
 Pred. No.: 6,98e-116 Length: 4299
 Score: 1717.00 Matches: 332
 Percent Similarity: 70.02% Conservative: 93
 Best Local Similarity: 54.70% Mismatches: 166
 Query Match: 51.19% Indels: 16
 DB: 6 Gaps: 9
 US-10-757-093-4 (1-634) x AX025617 (1-4299)
 QY 33 HisGluGlnProLeuIleLysValArgProGlnArgThrSerSerArgGluLeuValAsn 52
 DB 2103 CAGCGCTGAGTGGCCACCATGTCCTCTGTAGAAACCCCAACCCGTGAATCAAAAAA 2162
 QY 53 LeuAspGlyLeuTyrLysPheAlaLeu-----AlaSerGlyLeuAsnAspThrAla 69

DB 2163 CTCGACGCGCTGTGGCATTCAGTCTGATGCGCAAAACGTGGAAATGATCAG----- 2216
 QY 70 GlnProTyrThrAlaProLeuProLysGlyLeuGluCysProValProAlaSerTyrAsn 89
 DB 2217 CGTTGGTGGGAAAGCGCCCTTACAGAAAGCCGGCAATTCCTGTGCGCAGGAGTTTAAAC 2276
 QY 90 AspIlePheIleSerArgGluIleHisAspHisValGlyTyrValTyrTyrGlnArgGlu 109
 DB 2277 GATCAGTTCGCCGATGCAATATTTGTATATATGCGGGCAACGTGTGATACGGCCGA 2336
 QY 110 ValIleValProLysGlyTyrSerGlnLysGlyTyrLeuValArgAlaGluSerAlaThr 129
 DB 2337 GTCTTATATACGAAAGTGTGGCAGCCAGCGTATGCTGTGCGTTTCGATGCGGTACT 2396
 QY 130 HisHisGlyArgIleTyrValAlaAsnAspArgLeuValAlaGluHisValGlyTyrThr 149
 DB 2397 CATTCAGCGCAAGTGTGGTCAATATATCAAGAAAGATGAGACATCAGAGGCGGTATACG 2456
 QY 150 ProPheGluAlaAspValThrGlnLeuValAlaProGlyGluLysPheArgLeuThrIle 169
 DB 2457 CATTTGAAGCCGATGTACAGCCGATGTATTGCGGGGAAAGTACGATATACCGGT 2516
 QY 170 GlyValAlaAsnAspGluLeuThrHisGluThrIleProProGlyLysIleThrThrGlyAsn 189
 DB 2517 TGTGTGAACAAAGACTGAATGCGACATACCGCGCGGAGATGTGATTAC--GAC 2573
 QY 190 AlaThrGlyLysArgIleGlnThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAla 209
 DB 2574 GAAACCGCAAAAGAGAGCTTACTTCATGATATCTTAACTTAATGCGGAATTCAT 2633
 QY 210 ArgSerIleTyrLeuTyrSerValProGlnGlnHisIleGlnAspIleThrValValThr 229
 DB 2634 CCGACGCTAATGCTTACACACGCGCAACACCTGGGAGAGATATACCGTGTGACG 2693
 QY 230 AspVal-----AspGlyAspAsnGlyLeuIleAsnTyrGluValGluValAlaAsnGln 247
 DB 2694 CATGTCGCGCAAGACTGTACACACCGCTGTGTGATGCTGACAGGTG--GTGCCAAT-- 2747
 QY 248 ThrThrGlyGlnIleGlnIleSerValIleAspGlyAlaIleValAlaLysAla 267
 DB 2748 -----GGGATGATGACGCTTAACCTGCTGATGCGGATCAACAGGTGTTCACTGA 2801
 QY 268 SerGlyAlaGlnGlyThrValThrIleProSerValLysLeuTyrGlnProGlyAlaAla 287
 DB 2802 CAAGGACATAGGGGAGCTTGGCAAGTGGAATCCGACCTTGGCAACCGGGTGAAGCT 2861
 QY 288 TyrLeuTyrGlnLeuGlnValAsnIleValGlySerSerGlyAspValAlaAspThrTyr 307
 DB 2862 TATCTCTATGAACGTGCGCTC-----ACAGCCAAAGCCAGACAGAGTGTATATCTAC 2915
 QY 308 AsnLeuAlaThrGlyValArgThrValLysValAlaGlySerGlnPheLeuIleAsnGly 327
 DB 2916 CCGCTTCGCGTGGGATCCGCTGACGTGAGCGATGAAAGGGAACAAGTCTGATTAACAC 2975
 QY 328 LysProPheTyrPheThrGlyPheGlyLysHisGluAspThrAlaValArgGlyLysGly 347
 DB 2976 AAACCGTTACTTACTGCTGCTTGGTGTGATGAAGATCCGACATTTGGTGCAAAAGGA 3035
 QY 348 HisAspProAlaTyrMetValHisAspPheGlnLeuMetLysTyrIleGlyAlaAsnSer 367
 DB 3036 TTGGATTAACGTGCTGATGTGACAGCAACGATTAATGAGATGAGTTGGGCAACTGC 3095
 QY 368 PheArgThrSerHisTyrProTyrAlaGluGluValMetAspPheAlaAspArgAsnGly 387
 DB 3096 TACCGTACTCGCATTAACCTTACGCTGAAGAGATCTGATCGGCAATGAATGATGCG 3155
 QY 388 IleValValIleAspGluThrProAlaValGlyLeuAsnIleAlaLeu---MetGlyVal 406
 DB 3156 ATCGGTGATATGATGAACCTGCTGTGCGCTTAACTCTTTAGCAATGTGTTTC 3215
 QY 407 SerGluSerGlyAlaPro---GlnThrPheThrProAspAlaIleAsnAspLysThrGln 425

Db	3216	GAAGGGGCGCAACAGCCGAAAGAAATGTGACAGCGAAGGCGAGCTCAACGGGGAATCTCAG	3275
Qy	426	GLUAAHLSLSYGLNALLILEARGLULEULEIALAAGABRYVAENHIEALSERVAL	445
Db	3276	CAACGGCACTTAACAGGCGATTAAAGAGGTGTATAGCGCTGTCAAAAACCCAGCCCTG	3339
Qy	446	VALMETTPSERILEALASNGLUPROLASERHISGLIUAERGLVALAARGLUYRPH	465
Db	3336	GTGATGTGAGATTATGTCACAGAACCCGGAATACCCCTCCGCAAGGTGCACGGGAATATTTC	3399
Qy	466	GLUPROLEUTHRASNLEUTHRARGGLINLEUABPROTHRARGPROLIETHRPHENLAASN	485
Db	3396	GGCGACATCGGGGGAAGCAACCGGTAACTCCAGCCGACGGCTCGATCACTCCGCTCAT	3455
Qy	486	VALGLYTHRALATHRYGLINLEUABRARGILESERABRLEUPHEABRVALSERCYSTLE	505
Db	3456	GTAAATGTTCTCGACGCTCACACCCATACCATCAGCGATCTCTTGATGTCTGTGCTCTG	3519
Qy	506	ASNRATGTYRPHESGLYTRPYRSEGLINHRGILYARPLEUGLUALAGLUALALALEU	525
Db	3516	AACCGTATTTACGGATGGTATGTCCAAAGCGCGATTTGAAACGGCAGAGAAGTACTG	3579
Qy	526	GLUYSGLIUEAHISGLYTRPGINLIUYSPHENIARPROLIEVALIMETHRGLUYR	545
Db	3576	GAAGAAAGAACTTCTGCGCTCGCAGAGAAATCGCATACGCCATTATCATCACCGAATAC	3635
Qy	546	GLYLAARPHRLEUALLGLYLEUHISSERILEULEUGLYLEUPROTPRSEGLIUDHE	565
Db	3636	GGCGTGATACGTTAGCCGGGCTGCATCATATGCACCGACATGTGAGATGAAGATAT	3695
Qy	566	GLINVALGIMELLEUABRMEITYNIAHARGVALPHEADPARGILEGLUSERMECHALACTY	585
Db	3696	CAGGTGCGATGAGCTGGATATGATATCACCGCGCTTTGATCGCGTCAAGCGCGTGTCTG	3755
Qy	586	GLUHISVALITTPRASNPHENLAASRPHESGLINHRASNLEUGLYLEIARGLVALASRGLY	605
Db	3756	GAACGAGTATGGAATTTGCGCGATTTTGCAGCTCGCAAGCAATTTGCGCTTGCGCGGT	3815
Qy	606	ASNLVSLYGLVALPHETHRARGNARGLYSPROLYSALALALANHSERLEUARG	625
Db	3816	AACAAGAAAGGATCTTCATCTCGACCGCAACCGAAGTGGCGGCTTTCTGTGTGCA	3879
Qy	626	ALATGTTPTPHRSERILEASR	632
Db	3876	AAACGCTGACGTGCATGAC	3896
RESULT 10			
ATH298139			
LOCUS			
DEFINITION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
SOURCE			

[illegible]

```

OY 215 TYSerValProGlnGlnHisIleGlnAspIleThrValValThrAspVal-----Asp 232
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 614 TACACACCCCGAAGACCTGGGGTGAACGATATCACCGTGTGATCGCATCGCGAAGAC 673
OY 233 GLYAspAsnGlyLeuIleAsnTyrGluValGluValAlaAsnGlnThrGlyGlnIle 252
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 674 TGTAAACACCGCTCTTGTGACTGGCAGGTG--GTGGCCAAAT-----GGTGAATGTC 721
OY 253 GlnIleSerValIleLeuAspGluAspGlyAlaIleValAlaIleValAsnSerGlyAlaGlnGly 272
    ::::: ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 722 AGCGTTGAAGTGGCTGATCGGATTCACAGCTGTTGCAACAGCAAGCAAGCAAGCGGG 781
OY 273 ThrValThrIleProSerValLysLeuTyrPglInProGlyAlaAlaTyrLeuTyrGlnLeu 292
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 782 ACTTTCAGAGTGTGAATCGGCACCTCGGCAACCGGGTGAAGTTATCTTAATGAACTCG 841
OY 293 GlnValAlaAsnIleValGlySerSerGlyAspValValAspThrThrAsnLeuAlaThrGly 312
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 842 TGGGTC-----ACAGCCAAAGCCAGACAGAGCTGATATCTTACCCCGCTTCGGCTCGC 895
OY 313 ValArgThrValLysValAlaGlySerGlnPheLeuIleAsnGlyLysProPheTyrPhe 332
    ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 896 ATCCGCTCAGTGGCAGTGAAGGCGAAGCAAGTTCTGATTAAACCAACCGCTTCTACTT 955
OY 333 ThrGlyPheGlyLysHisGluAspThrAlaValArgGlyLysGlyHisAspProAlaTyr 352
    ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 956 ACTGGCTTGTGTGTCATGAAGATGCGGACTTACGTGGCAAGAGATTGCTGAACGTCGTCG 1015
OY 353 MetValHisAspPheGlnLeuMetLysTyrIleGlyAlaAsnSerPheArgThrSerHis 372
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 1016 ATGTGTCAGACACCGCATTAATGATGCAATGGGGCCAACTCTTAACCTGACTGCGAT 1075
OY 373 TyrProTyrAlaGluGluValaMetAspPheAlaAspArgAsnGlyIleValIleAsp 392
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 1076 TACCTTAACCTGAGATGATGCTGCACTGGGACGATGAACATGACATCGGTGATGAT 1135
OY 393 GluThrProAlaValGlyLeuAsnIleAlaLeu---MetGlyValSerGlyLysValAla 411
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 1136 GAAACTCTCTGCTGGCTTTAACTCTCTTAACTCTTAACTCTTAACTCTTAACTCTTAACTCT 441
OY 412 Pro---GlnThrPheThrProAspAlaIleAsnAspLysThrGlnGluAlaHisLysGln 430
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 1196 CCGAAAGAACTGTACAGCAAGAGCAAGCAAGTCAACGGGAAACTCAAGCGCACTTACAG 1255
OY 431 AlaIleArgGluLeuIleAlaArgAspLysAsnHisAlaSerValValMetTyrPheIle 450
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 1256 GCGATTAAAGAGTGAATGCGCGTGAACAAAACCAACCGCGTGAATGTGAATGAT 1315
OY 451 AlaAsnGluProAlaSerHisGluAspGlyValaArgGluTyrPheGluProLeuThrAsn 470
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 1316 GCCAACGAAACCGGATACCGCTCCGCAAGGTGACGGAATATTTTCGGCCACATGGCGGAA 1375
OY 471 LeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGlyThrAlaThr 490
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 1376 GCAACGCTTAACCTGACCGACCGGCTCGATCACTCGCGCAATGTAATGTTCTGCGAC 1435
OY 491 TyrGlnLeuAspArgIleSerAspLeuPheAspValaSerCysIleAsnArgTyrPheGly 510
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 1436 GCTGACACCGGATACGATGATCTTTATGATGCTGTCGTCGAACCGTTATTAACGGA 1495
OY 511 TrpTyrSerGlnThrGlyAspLeuGluGluValaGlnAlaIleLeuGluLysGlnLeuHis 530
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 1496 TGGTATGTCCAAAGCGCGCATTTTGAACCGCGCAAGAAAGTACTGAAAAAGAACTTCTG 1555
OY 531 GlyTyrGlnGluLysPheHisArgProIleValMetThrGluTyrGlyValaAspThrLeu 550
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 1556 GCGTGGGAGGAAAGTCACTGACCGCATTAATCAACCGAATGCGCGTGAATACGTTA 1615
OY 551 AlaGlyLeuHisSerIleLeuGlyLeuProTyrPheSerGluGluPheGlnValGlnMetLeu 570
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 1616 GCCGGGTCGCACTCAATGATACACCGACATGTGAGTGAAGATACAGTGTGATGGCTG 1675

```

```

OY 571 AspMetTyrHisArgValPheAspArgIleGluSerMetAlaGlyGluHisValTyrPheAsn 590
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 1676 GATATGATATCACCGGCTCTTGTGATCGCTCAAGCCCGTCGTGGTGAACAGGTATGGAAT 1735
OY 591 PheAlaAspPheGlnThrAsnLeuGlyIleIleArgValaAspGlyAsnLysGlyVala 610
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 1736 TTCGCCGATTTTGGCAGCTCCGCAAGGCAATATGCGCGTGGCGGTAAACMAAAGGGATC 1795
OY 611 PheThrArgAspArgLysProLysGluAlaAlaHisSerLeuAsnGlyAlaArgTyrThrSer 630
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 1796 TTCCTCCGACCGCAACCGAAGTGGCGGCTTTTCTGCTGCAAAAACGCTGACATCGC 1855
OY 631 IleAsp 632
    :::::
Db 1856 ATGAAC 1861

RESULT 11
ATH298137 11509 bp DNA linear PLN 25-MAR-2003
LOCUS ATH298137
DEFINITION
  Arabidopsis thaliana eIF-4A1 gene for translation initiation factor
  eIF-4A1, exons 1-5 and gus gene for beta-glucuronidase.
ACCESSION
  AJ298137
VERSION
  AJ298137.1 GI:14594801
KEYWORDS
  beta-glucuronidase; eIF-4A1 gene; GUS gene; translation initiation
  factor eIF-4A1.
SOURCE
  Arabidopsis thaliana (thale cress)
  Arabidopsis thaliana
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
  rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
  1 De Greve, H., Nguyen, V., Deboeck, F., Thi-Doong, L., Karimi, M. and
  Hernalsteens, J.P.
  T-DNA tagging of the translation initiation factor eIF-4A1 from
  Arabidopsis thaliana
  Plant Sci. 161, 685-693 (2001)
  2 (bases 1 to 11509)
  de Greve, H.
  Direct Submission
  Submitted (18-DEC-2000) De Greve H., Biologie, Vrije Universiteit
  Brussel, Genetische Virologie, Vrije Universiteit Brussel,
  Paardenstraat 65, Sint Genesius Rode, B-1640, BELGIUM
FEATURES
    source
        1..11509
            /organism="Arabidopsis thaliana"
            /mol_type="genomic DNA"
            /strain="C24"
            /db_xref="taxon:3702"
            /chromosome="3"
            /note="transgenic line M18"
        944..3532
            /gene="eIF-4A1"
            /join(944..1068,1717..1785,1817..1973,2064..2494,
            2576..3532)
        944..1068
            /gene="eIF-4A1"
            /number=1
            /gene="eIF-4A1"
            /number=1
        1069..1716
            /gene="eIF-4A1"
            /number=1
        1717..1785
            /gene="eIF-4A1"
            /number=2
        1786..1816
            /gene="eIF-4A1"
            /number=2
        1817..1973
            /gene="eIF-4A1"
            /number=3
            /join(1931..1973,2064..2494,2576..3211)
            /gene="eIF-4A1"
            /codon_start=1

```

```

/product="translation initiation factor eIF-4A1"
/protein_id="CAC43286.1"
/db_xref="GI:14594802"
/db_xref="GOA:Q94CM1"
/db_xref="UniProt/TREMBL:Q94CM1"
/translation="MGLQENLRLGIYAFGEKPSAIOQGIYVFCGLDVIQQAQSGT
GVTATGSGVLIQQLDPSLIQCALVLAFLRELQAQIEKMRALADYLVKVAHCVGCT
SVREDRITLQAGHYVVGTRGVFMDLKRQSLRADIIMKPVLEADEMILSRGKDOIY
DIFQLLPKIQGVSPATMPREALETTRKMSKPVATILYKREDTIEGIKQPVNVEK
EEMKLETTCDLVEYLAITQSIVFVNTNRKVDLTDMRSRDHTVSTHDMQNTDI
IKREFSGSRVLIITDILARQIDVOQSLVINFDLPQPEYLARIIGSGRGKGV
AINEFTRDDERMIFDIQEFYVVEELPSNVADLL"
1974..2063
/gene="eIF-4A1"
/number=3
2064..2494
/gene="eIF-4A1"
/number=4
2495..2575
/gene="eIF-4A1"
/number=4
2576..3532
/gene="eIF-4A1"
/number=5
3533..5683
/gene="gus"
3533..5683
/gene="gus"
3533..5683
/codon_start=1
/product="beta-glucuronidase"
/protein_id="CAC43287.1"
/db_xref="GI:14594803"
/db_xref="GOA:Q93VY4"
/db_xref="UniProt/TREMBL:Q93VY4"
/translation="MVRPVEIPREIKKLDGLMAFSLDRENCIDORWESALQESRA
IAPGSPNQFADADIRNAYAGVYQREVEIPKMGAGORIVLFPDAVTYGVKVVNO
EWMHQGGYTPFADVPYVIAKSVRTVCVNSLNMWOTIPGAVITBENGKKQSY
FHFERNYAGIHSVNLKTTNPVVDITVTVYADONASVDMOVAVADSVLRAD
ADQVAVATGCGTSGTLQVNPVPHLMQPBEGVLYELCTAASQTECDIYPLRIGTRSYA
KGEQFLINRPFYFTGFRHEDADLRKGFNDVLMYDHALMDWIGANSRTRSHYVA
KEMLDADDEHGIIVIDETAVGFNLSLGIFEGANPKPELISEAVNGFTQVNLQAI
KEMLDADKNSPSVVMMSIANEPDTRPOGAEFAPLAELATRKLDPPRTRCOVNMFC
AHTDITSDLEFVLCINRYGMYVQSGDLTAELKELLMQEKHOPITITEYGV
TLAGHSMTYDMSSEYQCALMDMYRVRVPRVSAVVGQVWNTADPAITSQGIIRVGN
KGIIFTRKPSAFLDQKMTGMFGKPKQGGKQ"

```

ORIGIN

Alignment Scores:

	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-10-757-093-4 (1-634) x ATH298137 (1-11509)	2,986-115	1716.00	70.43%	54.98%	11509	331	93	162	16	9

```

115 GLYTRPserGlnGluArgTyrLeuValArgAlaGluSerAlaThrHisGlyArgIle 134
1103 GGTGGGACAGCCAGCGTATCGCGCTTTCGATCGCGTATCATTAACGGCAAGG 4162
135 TyrValAsnAspArgLeuValAlaGluHisValGlyTyrThrProPheGluAlaAsp 154
4163 TGGGATTAATATCAGGAAGTATGAGCATCAGGGCGGCTATCGCCATTGAAAGCGGAT 4222
155 ValThrGluLeuValAlaProGlyGluValPheArgLeuThrIleValAsnAsnGlu 174
4223 GTCAACCCGATGATATTCGCCGGAAGATGATACGATACCTTTGTGGAACAACGAA 4282
175 LeuThrHisGluThrIleProPheGlyValIleThrThrGlyAsnAlaThrGlyValArg 194
4283 CTGAACCTGACAGCTATCCCGCGGAAATGATGATTAAC---GACGAAAACGGCAAGAA 4339
195 IleGlnThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaArgSerIleThrPleu 214
4340 AAGCAGTCTTACTTCCATGATTTCTTAATCATGCGGAATCCATCGCAGCGTAATGCTC 4399
215 TyrSerValProGlnGlnHisIleGlnAspIleThrValValThrAspVal-----Asp 232
4400 TACACCCAGCCGCAACCTGGGTGAGCATATCACCGTGTGACGATGCGCGCAAGAC 4459
233 GlyAspAsnGlyLeuIleAsnTyrGluValGluValAlaAsnGlnThrThrGlyGlnIle 252
4460 TGTACACAGCGCTGTGATCGGACGCTG---GTGACCAAT-----GGTGAATGC 4507
253 GlnIleSerValIleAspGluAspGlyAlaIleValAlaValAsnAspGlyAlaGlnGly 272
4508 AAGCTTGAACCTGGTATCGGATCAACAGTGTGTGCAACATGACCAAGGACATACCGGG 4567
273 ThrValThrIleProSerValIleValLeuThrGlnProGlyAlaAlaTyrLeuTyrGlnLeu 292
4568 ACTTTCAGATGGTGTGATCCGACCTCTGCGCAACCGGATGATATCTATGAACTG 4627
293 GlnValAsnIleValGlySerSerGlyAspValValAspThrTyrAsnLeuAlaThrGly 312
4628 TGGCTC-----ACAGCCAAAGCCAGACAGATGATATCTACCCGCTTCGCGTCGCG 4681
313 ValArgThrValIleValAlaGlySerGlnPheLeuIleAsnGlyIleAspPropheTyrPhe 332
4682 ATCCGCTCAGTGCAGTGAAGGCGCAACGATCTCTGATTAACCAACAACGCTCTACTTT 4741
333 ThrGlyPheGlyIleHisGluLeuAspThrAlaValArgGlyIleGlyIleAspPropheAlaTyr 352
4742 ACTGCTTTGCTGCTGATGATGCGAATGCGAATGCGAATGCGAATGCGAATGCGAATGCG 4801
353 MetValHisAspPheGlnLeuMetCysTyrPheIleGlyAlaAsnSerPheArgThrSerHis 372
4802 ATGCTCAGACGACGACGATTAATGAGCTGATGGGCGCAACTCTTACCTGATCTGCAT 4861
373 TyrProTyrAlaGluGluValMetAspPheAlaAspArgAsnGlyIleValValIleAsp 392
4862 TACCTTACGCTGAAGATGCTGCACTGGCGAGATGAAATGATGATGATGATGATGATGAT 4921
393 GlnThrProAlaValIleLeuAsnIleAlaLeu---MetGlyValSerGluSerGlyAla 411
4922 GAAATGCTGCTGCTGCTGCTTAACTCTCTTGAAGATGGTTCGAAAGCGGCAACAG 4981
412 Pro---GlnThrPheThrProAspAlaIleAsnAspIleThrGlnGluAlaHisIleGln 430
4982 CCGAAGAACTGTACAGCGAAGGAGGAGTCAACGGGAACTGACGACGCACTTACAG 5041
431 AlaIleArgGluLeuIleAlaArgAspIleAsnHisAlaSerValIleMetTyrSerIle 450
5042 GCGATTAAAGAGTATAGGCGGTGACAAACCAACCAAGCGTGGATGATGATGATGATGAT 5101
451 AlaAsnGluProAlaSerHisGluAspGlyAlaArgGlyIleThrPheGluProLeuThrAsn 470
5102 GCCAAGCAACCGGATCCGCTCCGCAAGGTGACCGGAAATATTTCCGCGCACGCGGCA 5161
471 LeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGlyThrAlaThr 490

```



```

Db      1154 CGTTGGATGCGGTCATCTATTACGGCAAGTGGGTCAATATACGAAAGATGAGAG 1213
Qy      144 HlvalaGlyGlyThrProheGluAAspValThrlGluValAlaProGlyGlu 163
Db      1214 CATCAGGGGGCGGCTATACGCATTGGAAGCCGATGTACAGCCGATGATTATGCGGGAA 1273
Qy      164 LysPheArgLeuThrIleGlyValaAsnGluLeuThrThrlleGluThrIleProProGly 183
Db      1274 AGGTACGATATCCGCTTGTGTGTGAACACGAACTGAACCTGGACATCCCGGGGA 1333
Qy      184 LysIleThrThrGlyAsnAlaThrGlyLysArgIleGlnThrThrlleAspPheTy 203
Db      1334 ATGGTATTAAC--GACGAAACCGCAAGAAAGAGTCTTACTTCCATGATTTCTT 1390
Qy      204 AsnTyrAlaGlyLeuAlaArgSerIleTyrPheTyrSerValProGlnGlnIleGln 223
Db      1391 AACTATGCGCGAATCATCGACGCTAATGCTCTACACCAACCGCGAAACCTGGTGAC 1450
Qy      224 AsnIleThrValValThrAspVal-----AspGlyAspAsnGlyLeuIleAsnTyrGlu 241
Db      1451 GATATACCGGTGTGACGATGTCGGCAAGACTGTAAACCAACCGCTGTGACGCGAG 1510
Qy      242 ValGluValAlaAsnGlnThrThrlleGlnIleGlnIleSerValIleAspGluAspGly 261
Db      1511 GTG--GTGGCCAT-----GGTGAATGCAAGGTTGAACCTGGGTGATGCGAATCA 1558
Qy      262 AlaIleValAlaLysAlaSerGlyAlaGlnGlyThrValThrIleProSerValLysLeu 281
Db      1559 CAGGTGGTGCACCTGACCAAGGCACTAGCGGGACTTTCGACATGGTGAATCCGCACTC 1618
Qy      282 TTPGlnProGlyAlaAlaTyrLeuTyrGlnIleGlnValaAsnIleValGlySerSerGly 301
Db      1619 TGGCAACCGGGTAAAGTTATCTTAAAGACTGTGCTC-----ACAGCCAAAGCGAG 1672
Qy      302 AspValAlaAspThrThrlleAsnLeuAlaThrGlyValaArgThrlleValAlaGlySer 321
Db      1673 ACAGACTGTGATCTACACCGCTTCGGTCGCGACATCGGCAATGGCAAGAGGCGAA 1732
Qy      322 GlnPheLeuIleAsnGlyLysProPheTyrPheThrGlyPheGlyLysIleAspThr 341
Db      1733 CAGTTCTGATTAACCAACAAACGCTTCTACTTACTGCTTGTGCTGATGAAGATGCG 1792
Qy      342 AlaValArgGlyLysGlyHisAspProAlaTyrMetValHisAspPheGlnLeuMetLys 361
Db      1793 GACTTCGCGGCAAGAGATTGATAACGTGCTATGCTGACGACCAACGATTAATGAC 1852
Qy      362 TPPIleGlyAlaAsnSerPheArgThrSerHisTyrProTyrAlaGluGluValaMetAsp 381
Db      1853 TGGATTGGGGCCAACTCTACCTTACTGCACTTACCCTTACCTGAAGAGATGCTGAC 1912
Qy      382 PheAlaAspArgAsnGlyIleValIleAspGluThrProAlaValGlyLeuAsnIle 401
Db      1913 TGGGCGAGATGAACATGATCGTGTGATGATGAACCTCTGCGGCTTAACTCTC 1972
Qy      402 AlaLeu---MetGlyValSerGlySerGlyAlaPro---GlnThrPheThrProAspAla 419
Db      1973 TCTTAAAGCATTTGTTTGAAGCGGCGCAACAGCCAAAGAACTGTACACGGAAGGCA 2032
Qy      420 IleAsnAspLysThrGlnGluAlaHisLysGlnAlaIleArgGluLeuIleAlaArgAsp 439
Db      2033 GTCAACGGGAAACTCAGCAAGCGCACTTACAGCGATTAAAGCTGAAGCGCGGAC 2092
Qy      440 LysAsnHisAlaSerValValMetTyrSerIleAlaAsnGluProAlaSerHisGluAsp 459
Db      2093 AAAAACACCCCAACGCTGTGATGTGATGATGATGATGATGATGATGATGATGATGAT 2152
Qy      460 GlyAlaArgGlyLysPheGluProLeuThrAsnLeuThrArgGlnLeuAspProThrArg 479
Db      2153 GGTGCGAGGGAATATTTCCGCGCACTGCGGAAAGCAACGCGTAACCTGACCGGACGCT 2212
Qy      480 ProIleThrPheAlaAsnValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeu 499

```

```

Db      2213 CCGATCACTGCGTCAATGATGTTCTGCGACGCTCACACCGATATCAAGCGATCTC 2272
Qy      500 PheAspValSerCysIleAsnArgTyrPheGlyTyrThrTyrSerGlnThrGlyAspLeuGlu 519
Db      2273 TTTGATGTGTGCTGCTGAACCGTTATTAAGATGTGATGTGTAAGCGCGGATTTGGA 2332
Qy      520 GluAlaGluAlaIleLeuGluLysGlyLeuHisGlyTyrGlnGluLysPheHisArgPro 539
Db      2333 ACCGCAAGAAAGTACTGAAADAAAGATTTCTGCGCTGCGAGGAAAGTCAATCGAC 2392
Qy      540 IleAlaMetThrGluTyrGlyAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeu 559
Db      2393 ATATCATCAACCGAATACCGCGTGAATGATGATGATGATGATGATGATGATGATGATGAT 2452
Qy      560 ProTyrSerGluGluPheGlnValaGlnMetLeuAspMetTyrHisArgValaPheAspArg 579
Db      2453 ATGTGAGTGAAGATGATAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2512
Qy      580 IleGluSerMetAlaGlyGlnHisValaTyrAsnPheAlaAspPheGlnThrAsnLeuGly 599
Db      2513 GTACAGCGCGCTGTCTCGTGAACAGTATGATGATGATGATGATGATGATGATGATGATGAT 2572
Qy      600 IleIleArgValaAspGlyAsnLysGlyValaPheThrArgAspArgLysProLysAla 619
Db      2573 ATATGCGCGTGGCGGTAACAAAGAGATCTTCACTCGCGACCGCAAAACCGAAGTGC 2632
Qy      620 AlaAlaHisSerLeuArgAlaArgTyrThrSerIleAsp 632
Db      2633 GCGGCTTTTCTGCTGCACAAAACGCTGACCTGATGATGATGATGATGATGATGATGATGAT 2671

RESULT 13
EC0414112      5667 bp      DNA      circular      SYN 27-SEP-2002
LOCUS          Cloning vector pGV1025 DNA, complete sequence.
DEFINITION     AJ414112
ACCESSION      AJ414112
VERSION        AJ414112.1 GI:23392741
KEYWORDS       Cloning vector pGV1025
SOURCE         Cloning vector pGV1025
ORGANISM       Other sequences; artificial sequences; vectors.
REFERENCE      1 Broeders,S., De Greve,H., De Rycke,R., Bauw,G., Jansens,S.,
                Buys,L., Smaghe,G., Ripoll,C., Peumans,W.J., Van Damme,E.J.M. and
                Herlaers,J.P.
                Leaf and bulb lectins from garlic (Allium sativum L.) are
                expressed, correctly processed and targeted to the cytoplasm in
                transgenic tobacco plants
                Unpublished
                2 (bases 1 to 5667)
                De Greve,H.
                Direct Submision
                Submitted (27-SEP-2001) De Greve H., Genetische Virologie, Vrije
                Universiteit Brussel, Paardenestraat 65, Sint-Genesius-Rode, B-1640,
                BELGIUM

FEATURES
     source      1..5667
                 /organism="Cloning vector pGV1025"
                 /mol_type="other DNA"
                 /specific_host="Escherichia coli"
                 /db_xref="taxon:176844"
     misc_feature 1..403
                 /note="pUC18"
     promoter     404..1263
                 /note="Caulliflower Mosaic Virus 35S promoter"
     gene         1294..3105
                 /gene="gus"
                 1294..3105
                 /gene="gus"
     CDS          1294..3105
                 /codon_start=1
                 /transl_table=11
                 /product="beta-glucuronidase"
                 /protein_id="CAD10672.1"
                 /db_xref="GI:23392742"

```


Oy	600	lelleaagvllaapglayashlytgcylvalphetrtrrargaaarglyvprolyala	619
Db	2971	ATATTGCGCGTTGGCGGTAAACAAGAAAGGATCTTCACTCGGACCGCAACCGAAGTC	3030
Oy	620	AlaAlaHisSerLeuArgAlaArgTyrHisSerIleaeP	632
Db	3031	GCGGCTTTTCTGCTGCACAAAACGCTGCAGCTGCATGAAC	3069
RESULT 14			
LOCUS	AY456412	14701 bp	DNA
DEFINITION	Cloning vector pZGA22, complete sequence.		
ACCESSION	AY456412		
VERSION	AY456412.1	GI:38327505	
KEYWORDS			
ORGANISM	.		
REFERENCE	Cloning vector pZGA22		
AUTHORS	Cloning vector pZGA22		
TITLE	other sequences; artificial sequences; vectors.		
JOURNAL	1 (bases 1 to 14701)		
AUTHORS	Mitterbauer, R., Poppenberger, B., Radtschning, A., Lucshyn, D.,		
TITLE	Lemmens, M., Gloessel, J., and Adam, G.		
AUTHORS	Toxin-dependent utilization of engineered ribosomal protein L3		
TITLE	limits trichothecene resistance in transgenic plants		
AUTHORS	Unpublished		
TITLE	2 (bases 1 to 14701)		
AUTHORS	Adam, G.		
TITLE	Direct Submission		
AUTHORS	Submitted (04-NOV-2003) Center of Applied Genetics, BOKU,		
TITLE	University of Natural Resources and Applied Life Sciences,		
AUTHORS	Muehgaese 18, Vienna A-1180, Austria		
FEATURES			
SOURCE	Location/Qualifiers		
	1..14701		
	/organism="Cloning vector pZGA22"		
	/mol_type="other DNA"		
	/db_xref="taxon:254255"		
	complement(2454..2478)		
	/note="T-DNA right border"		
	2519..4277		
	/gene="nptII"		
	2519..2829		
	/gene="nptII"		
	2838..3632		
	/gene="nptII"		
	/codon_start=1		
	/product="neomycin phosphotransferase II"		
	/protein_id="AAK17784.1"		
	/db_xref="GI:38327506"		
	/translation="MIEQDGLHAGSPAAWVERLFGYDWAQQTICSDPAVRLASQGR		
	PVLKPTDLSGALNDEARLISWLTATPGPCAAVLDPVTEAGRDMLLGEVGODI		
	LSSHLAPAEKYSIMADNARLRHTLDPACPCGDAKRIEARTRMEGVLVDODDLE		
	EHQGLAPAELEPARLKAIRPDSDDLVYTHGDCLEINIVENKRSFGFIDCGRLGADRY		
	QDILATRDIAEEIGGEWADRFLVLYGIAAPDSRIAFYRLDF"		
	4022..4277		
	/gene="nptII"		
	4974..7924		
	/gene="gusA"		
	4974..5808		
	/gene="gusA"		
	/note="from CamV 35S"		
	5845..7656		
	/gene="gusA"		
	/codon_start=1		
	/product="beta-glucuronidase"		
	/protein_id="AAK17785.1"		
	/db_xref="GI:38327507"		
	/translation="MLRVEVTPTRRIKKLDGLMAFSLDRENGCIDORWMESALQESRA		
	IAYGSPFDQPADADIRNYAGNVVQREVFIPKMGAGQRIVLRPDAVTHYKRWNNQ		
	EVMEOGXYTPPEADVPYVYTAGKSVRTTCVNNELMNOITPPGVITIDENGGKSSY		
	FHDPEVYAGIRHSWLYTTPNTWDDITVYHVAQDNHASVDMQVANGDSVLELSD		
	ADQGVATGCGTSGTLOVYVPHLMQPSGTYIYELCYTKASGTEDCIYIPLRGITSVAV		
	KGEQFLINHRPFYFTGFRHSDADLRGQFDNVNLVHDHLMNDIGASTYRTSYPAA		
	EEMDMDWDEHGIIVVIDETAAAGFNISLGI GEADNKPVELYSEBAVNGETQAAHLQAI		

[illegible]


```

EEMLDWADHGIIVIDETIAVGFNLIGIFEAQNKPKELYSBEAVNGETOALQAI
KELIARDKHPVVMWSIANEPDTPQGAEPYALAEATKIDPRITCVAMFCD
AHTDITSDLEPDLNRYGVYVOSGLEAEKYLEKLEAKIDKHOPIIIEYGD
TLAGHSMTDMSEVYQACMLDMHYRVRPVSVAVGQVMNADPATISGLIRVGN
KKKIFTRDKRPSAAFLQKRTGMFGEPQGGKQ"
terminator
7727.7979
/misc_feature
/comp= "NOS"
/comp= (8621.08646)
/comp= "T-DNA left border"
/misc_feature
/comp= (9156.10198)
/comp= "similar to ceca in GenBank Accession Number L13842"
/misc_feature
/comp= (10199.11680)
/comp= "similar to cifa in GenBank Accession Number X00713"
/misc_feature
/comp= (11681.12673)
/comp= "similar to NPTIII gene in GenBank Accession Number V01547"
/misc_feature
/comp= (12674.13443)
/comp= "similar to transposable element IS1 in GenBank Accession Number X58999"
/misc_feature
/comp= (13444.13794)
/comp= "similar to NPT III gene in GenBank Accession Number V01547"
/misc_feature
/comp= (13795.14066)
/comp= "similar to kila in GenBank Accession Number M62846"
/misc_feature
/comp= (14141.14758)
/comp= "ori V: similar to sequence in GenBank Accession Number M20134"

```

ORIGIN

Alignment Scores:

```

Pred. No.: 4,47e-115 Length: 14758
Score: 1715.50 Matches: 338
Percent Similarity: 68.25% Conservative: 94
Best Local Similarity: 53.40% Mismatches: 182
Query Match: 51.15% Indels: 19
DB: 12 Gaps: 10

```

US-10-757-093-4 (1-634) x AF485783 (1-14758)

```

QY 10 LeuSerLeuAlaAlaProSerLeuGlyThrProAlaAlaArgHisPheProAlaGln 29
DB 5749 CTATCCTTCGCAACACCTTCCTCTATATAGGAAGTTCATTTCATGAGAAACACG 5808
QY 30 MetThrGlnHisGlnGlnPro-----LeuIleValAlaArgProGlnAgtThrSer 46
DB 5809 GGGGACTCTAGAGAGATCCCGGGGTGTCACTCCCTTATGTACGTCGTGAGAAACCCA 5868
QY 47 SerArgGlnLeuValAsnLeuAspGlyLeuTrpLysPheAlaLeu-----AlaSer 63
DB 5869 ACCCGGAAATCAAAAACCTCGACGGCTGTGGCATTCAGTCTGATCGCGAAACTGT 5928
QY 64 GlyLeuAsnAspThrAlaGlnProTrpThrAlaProLeuProLysGlyLeuGlnCysPro 83
DB 5929 GGAATGATCAG-----CGTTGGTGGGAAAGCGCGTTACCAAGCCGGCAATTCCT 5982
QY 84 ValProAlaSerLysAsnAspLysPheIleSerArgGlnIleHisAspHisValGlyTrp 103
DB 5983 GTGCCAGGCGAGTTTAAAGATTCGCGCATGCAATATTCGTAATTATGCGGGGCAAC 6042
QY 104 ValTrpGlnArgGlnValIleValProLysGlyTrpSerGlnGlnAgtGlyLeuVal 123
DB 6043 GTTCGTATACGGCGAAGCTTTATACCAAGAGTTGGGCAAGCCGATATCGTCGT 6102
QY 124 ArgAlaGlnSerAlaThrHisHisGlyArgGlyLeuValAsnAspArgLeuValAlaGln 143
DB 6103 CGTTTGGATCGCGTCACTCATTCAGCGCAAGTGTGGGTCAATATACGAAAGTATGAG 6162
QY 144 HisValGlnGlyTrpThrProPheGlnAlaAspValThrGlnLeuValAlaProGlyGlu 163
DB 6163 CATCAAGGGCGGTATACGCAATTTGAAAGCCGATGTCAAGCGGTATGATTCGCGGAAA 6222
QY 164 LysPheArgLeuThrIleGlyValAsnAspGlnLeuThrHisGlnTrpIleProGly 183
DB 183 LysPheArgLeuThrIleGlyValAsnAspGlnLeuThrHisGlnTrpIleProGly 183

```

```

DB 6223 AGTGAACGATATCACCCTTGTGTGAACAAGAACTGACGACATACCCCGCGGA 6282
QY 184 LysIleThrThrGlnValAsnAlaThrGlyLysArgIleGlnThrTrpGlnHisAspPheTrp 203
DB 6283 ATGTGATTTAC---GACGAAACCGCAGAGAAAAGACAGCTTACTTCCATGATTTCTTT 6339
QY 204 AsnTrpAlaGlyLeuAlaAspSerIleTrpLeuTrpSerValProGlnGlnHisIleGln 223
DB 6340 AACTATGCGCGAATTCATGCGACAGCGATGATCTTACACCAAGCCGAAACACTGGGTGAC 6399
QY 224 AspIleThrValValThrAspVal-----AspGlyAspAsnGlyLeuIleAsnTrpGlu 241
DB 6400 GATATCACCGGTGTGACGATGTGGCGCAAGACTGATACCAAGCCGCTGTGTGCTGGCAG 6459
QY 242 ValGlnValAlaAsnGlnThrGlnGlnIleGlnIleSerValIleAspGluAspGly 261
DB 6460 GTG---GTGGCCAT-----GGTATGTAGCGGTGAACCTGCGTGTGCGGATCAA 6507
QY 262 AlaIleValAlaLysAlaSerGlyAlaGlnGlyThrValThrIleProSerValLysLeu 281
DB 6508 CAGGTGTTCGACACTGACAGGACACTAGCGGACTTTCGAACTGGTGAATCCGCACTC 6567
QY 282 TrpGlnProGlyAlaAlaTrpLeuTrpGlnLeuGlnValAsnIleValGlySerGly 301
DB 6568 TGGCAACCGGGGTGAAGATTATCTCTATGAACGTGGCTC-----ACAGCCAAAGCCAG 6621
QY 302 AspValValAspThrTrpAsnLeuAlaThrGlyValArgThrValLysValAlaGlySer 321
DB 6622 ACAGAGTGTATATTCACCCGCTTCGCGGATCCGTCAGTGGCAGTGAAGAGGCGAA 6681
QY 322 GlnPheLeuIleAsnGlyLysProPheTrpPheThrGlyPheGlyLysHisGlnAspThr 341
DB 6682 CAGTTCTGATTAACACAAACCGCTTCTATCTAGCTGCTTGTGTGATGAAGATCGG 6741
QY 342 AlaValArgGlyLysGlyHisAspProAlaTrpMetValHisAspPheGlnLeuMetLys 361
DB 6742 GACTTCGTGGCAAGATTCGATTAAGCTGATGTGTCAGCAGCAGCATTAATGAGAC 6801
QY 362 TrpIleGlyAlaAsnSerPheArgThrSerHisTrpProTrpAlaGlnGluValMetAsp 381
DB 6802 TGGATTGGGGCCAACTTCACCGTACTGTCATTCCTTACCTTACCTTACCTTACCTTAC 6861
QY 382 PheAlaAspArgAsnGlyLysValIleAspGlyThrProAlaValGlyLeuAsnIle 401
DB 6862 TGGGCAATGACATGACATCGTGTGATGATGATGAACCTGCTGCTGCTTAACTTC 6921
QY 402 AlaLeu---MetGlyValSerGlnSerGlyAlaPro---GlnThrPheThrProAspAla 419
DB 6922 TCTTTAGGCAATTCGTTTCGAAGCGGCAACAAGCGAAAGAACTGATACGGAAGGCA 6981
QY 420 IleAsnAspLysThrGlnGlnAlaHisLysGlnAlaIleArgGlnLeuIleAlaGlyAsp 439
DB 6982 GTCAACGGGGAACCTGACAGAGCGCACTTACAGGCATTAAGAGCTGATAGCGCGTGAC 7041
QY 440 LysAsnHisAlaSerValValMetTrpSerIleAlaAsnGlnProAlaSerHisGluAsp 459
DB 7042 AAAAACCACCAAGCGGTGTGATGTGATGTGATGTGATGTGATGTGATGTGATGTGAT 7101
QY 460 GlyAlaArgGlnTrpPheGlnProLeuThrAsnLeuThrArgGlnLeuAspProThrArg 479
DB 7102 GTTGCACGGGAATATTCGCGCACTGGCGGAAGCAACGGGTAACTGACCCGAGCGGT 7161
QY 480 ProIleThrPheAlaAsnValGlyThrAlaThrTrpGlnLeuAspArgIleSerAspLeu 499
DB 7162 CCGATACCTGCGCTCATGTAATGTTCTGCGACGCTCAGACCCATACCATACCGATCTC 7221
QY 500 PheAspValSerCysIleAsnArgTrpPheGlyTrpTrpSerGlnThrGlnAspLeuGlu 519
DB 7222 TTGATGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7281
QY 520 GlnAlaGlnAlaLeuGlnLysGlnLeuHisGlyTrpGlnGlnLysPheHisAspPro 539
DB 7282 ACGGACAGAAAGTACTGAAAAAGAACTTCGCGCTGCGAGGAGAACTGCAAGCCG 7341

```

```
QY 540 ILeValMetThrGluTyrGlyAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeu 559
Db 7342 ATTATCATCACCAATACGGCGGATACGTTAGCCGGCTGCATCATGTACACCGAC 7401
QY 560 ProTyrSerGluGluPheGlnValGlnMetLeuAspMetTyrHisArgValPheAspArg 579
Db 7402 ATGTGAGTGAAGAGTATCATGTGCATGTGATGTATCATCACCGCTCTTGATCGC 7461
QY 580 IleGluSerMetAlaGlyGluHisValTyrAsnPheAlaAspPheGlnThrAsnLeuGly 599
Db 7462 GTCAGCGCGCTCGTCGATGAGAACGATGAAATTTGCCGATTTTGCACCTCGCAAGGC 7521
QY 600 IleIleArgValAspGlyAsnLysGlyValPheThrArgAspArgLysProLysAla 619
Db 7522 ATATTCCGCGCTGGCGGTACAGAAAGGATCTTCACTCGCGAACCGCAACCGAAGTCCG 7581
QY 620 AlaAlaHisSerLeuArgAlaArgTyrThrSerIleAsp 632
Db 7582 GCGGCTTTTCTGCTGCAGAAACGCTGACCTGCGCATGAAC 7620
```

Search completed: March 22, 2005, 20:24:58
Job time : 6487 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_p2n model

Run on: March 22, 2005, 18:23:41 ; Search time 4458 Seconds
(without alignments)
5413.357 Million cell updates/sec

Title: US-10-757-093-4
Perfect score: 3354
Sequence: 1 MKRLTGLSLSLAAPSIGTP.....RKPKAAHSIRARWTSIDKN 634

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame_p2n.model -DEV=xlp
-Q=/cgr2_1/USPRO.spool_p/US10757093/runat_18032005_164456_27770/app.query.fasta_1.775
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=bl0sum62 -TRANS=human40.cdi -LIST=45
-DOCALLGN=200 -THR SCORE=pcr -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pro -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10757093_QCGN_1_1_5180@runat_18032005_164456_27770 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: gb_esc1:*
2: gb_esc2:*
3: gb_hic:*
4: gb_esc3:*
5: gb_esc4:*
6: gb_esc5:*
7: gb_esc6:*
8: gb_ges81:*
9: gb_ges82:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1267	37.8	2473	3	AY321342 Rattus no
2	1253.5	37.4	2095	6	CD014094 CD0135027
3	1175	35.0	1995	6	CD014092 90134967
4	1168.5	34.8	2274	3	AK041058 Mus muscu
5	1045.5	31.2	1853	6	CD014093 90135266
6	946	28.2	1377	3	CRS93823 full-leng
7	803	23.9	1124	6	CD503076 CDA60-C07
8	797.5	23.8	1051	5	BX363460 BX363460
9	797	23.8	921	9	CL486845 SAIL_443

10	770.5	23.0	1138	6	CD503098
11	726.5	21.7	857	5	BX745933
12	712.5	21.2	811	9	CL679241 PR10125C
13	705.5	21.0	906	5	BQ941196
14	689	20.5	1055	4	BMS57676
15	683	20.4	877	7	CF406520
16	681.5	20.3	767	7	CF255373
17	674.5	20.3	740	7	CF521612
18	671.5	20.0	816	7	CK467358
19	671.5	20.0	878	7	CN155220
20	661.5	19.7	914	6	CB203472
21	656	19.6	1063	5	BX401772
22	649.5	19.4	915	5	BQ678153
23	648.5	19.3	846	6	CA453907
24	647.5	19.3	802	5	BU357212
25	643	19.2	733	7	CN264047
26	637	19.0	736	5	BP454056
27	633.5	18.9	791	5	BU239978
28	630.5	18.8	689	7	CR442804
29	622.5	18.6	725	5	BU315788
30	621	18.5	566	7	CV169266
31	620.5	18.5	900	6	CA489317
32	609.5	18.2	799	5	BU366987
33	609	18.2	771	4	BG121498
34	605.5	18.1	731	7	CV105482
35	601.5	17.9	780	7	CO559438
36	601	17.9	680	7	CK836522
37	591	17.6	898	5	BU179563
38	590	17.6	962	5	BQ684422
39	587.5	17.5	773	7	CF748482
40	585.5	17.5	757	4	B1851705
41	583.5	17.4	847	5	BX621253
42	582	17.4	925	5	BQ679162
43	578.5	17.2	624	4	B145636
44	577.5	17.2	673	6	CB841568
45	573	17.1	857	6	CB559443

ALIGNMENTS

RESULT 1
AY321342
LOCUS AY321342 2473 bp mRNA linear HTC 16-JUN-2003
DEFINITION Rattus norvegicus Ac2-223 mRNA, complete cds.
ACCESSION AY321342
VERSION AY321342.1 GI:32527744
KEYWORDS
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE
1 (bases 1 to 2473)
Xu,C.S., Li,W.O., Li,Y.C., Wang,L., Wang,S.F., Han,H.P., Wang,G.P.,
Chai,L.Q., Yuan,J.Y., Yang,K.J., Yan,H.M., Chang,C.F., Zhao,L.F.,
Ma,H., Shi,J.B., Rahman,S., Wang,Q.N. and Zhang,J.B.

TITLE
JOURNAL
Unpublished
2 (bases 1 to 2473)
XU,C.S., LI,W.O., LI,Y.C., WANG,L., WANG,S.F., HAN,H.P., WANG,G.P.,
CHAI,L.Q., YUAN,J.Y., YANG,K.J., YAN,H.M., CHANG,C.F., ZHAO,L.F.,
MA,H., SHI,J.B., RAHMAN,S., WANG,Q.N. and ZHANG,J.B.

TITLE
JOURNAL
Direct Submission
Submitted (11-JUN-2003) Henan Bioengineering Key Lab, Henan Normal
University, NO. 148 Jianshe Road, Xinxiang City, Henan 453002, P.R.
China

FEATURES
source
1. 2473
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
41. 2467
/note="liver regeneration related protein LRRC134"

```

/codon_start=1
/product="Ac2-223"
/protein_id="AAP86274.1"
/db_xref="GI:32527745"
/translation="MKRDKFQEGVEKYLMDLNSRVLAQNERDESSAKRPSWTAGI
PASMSPRSVCMFVLGQLCSALAQGMLEPKETPSRELKVLDDGMSFPADYNNR
LQGEKQWYRQPLRESGPTLMDPSSFNDDITQBAELNFGWVYREAEALPQRMTQ
DTRRVLRLINSAYYAVVWNGIHVHEGGHLPEFADITKLQOSPLTTPRTTIAI
NNTLTPYTLPGTIVYKTPSPMKPGYFVODISDFPMDAGLHSVLYTTPPTIID
ITVTTDVRDYGIVNWYISUGSDHFOLEFVLLDBDKIYAKGTNGSGQLEPRAHLM
KPYLMEHPALYLSLEVTMTTPSVSDFTLPVGIKTVAVTKSPFLINGKRFYQVNA
KHEBSDLRGRGFWPLIKDNLRLWLGANSFRSHQLPYSEEVYLCDDRGIIVIDEC
PGVGIYLPQSEFGNVLRLHLEVMDELVARDKHNPVVMWSVANSPLKAPAGYFEL
IVFOGLKTSDBSPRTGTVLWVYHAPVYDVCNVSLSWYDGHLEVILOLQSFEN
WYKMYOKPILIQSEYGADAVSGHEDPRMSESEYOTALLNHYHLILDEKRKYVIGEL
IMNFADPMTQNSPLRVTGNKKGIPTTRORNPMAAFIIRBYWIRIANTRGLIQVNSF
VVDDEKEVOHICGDADDAKLQHKVONVGDSTIAAGALICSTCTNRHAEKCKIST
VPSDDAQAQERGOQSLQATGMVAVTNKASHFQVROLKTTVHAHSIAVFKPDRPLNS
CLLDISAGSA"

ORIGIN

Alignment Scores:
Pred. No.: 1,66e-130 Length: 2473
Score: 1267.00 Matches: 266
Percent Similarity: 57.74% Conservative: 96
Best Local Similarity: 42.42% Mismatches: 193
Query Match: 37.78% Indels: 72
DB: 3 Gaps: 13

US-10-757-093-4 (1-634) x AY321342 (1-2473)

QY 42 ProGlnArgThrSerSerArgGluLeuValAsnLeuAspGlyLeuTrpLysPheAlaLeu 61
Db 269 CCGAAGAGAGACCCCTTCGCGGAACTCAAGAGTGTGGACGCGCTGTGAGACTTCCTG--- 325
QY 62 AlaSerGlyLeuAsnAspThrAlaGlnPro-----TrpThrAlaProLeu 76
Db 326 GCCCATTAATCTGAAACAATTCGGTTCGACAGGCTTCGAGAGACAGTGGTACCGGACCGCTA 385
QY 77 ProLysGly-----LeuGluCysProValProAlaSerTrpAsnAspIlePheIle 93
Db 386 CGGAGATCGGGGCCCACTTGACATGCGGTCCTTCGAGCTTCAATGACATACACCCAA 445
QY 94 SerArgGluIleHisAspHisValGlyTrpValTyTrpGlnArgGluValIleValPro 113
Db 446 GAAGCAAGAGCTTCGAACTTCATTTGGTGGTGTGTGATGAAGCGAAGAGAGCTTCCA 505
QY 114 LysGlyTrpSerGlnGlu-----ArgTrpLeuValArgAlaGluSerAlaThrHis 130
Db 506 CAGCGATGACCCAGGACCGGACAGAGAGTGTGTGTGATGAACTAAGAGAGGTCACTCC 565
QY 131 HisGlyArgIleTyValAsnAsnArgLeuValAlaGluHisValGlyTyTrpThrPro 150
Db 566 TATGCAATGTGTGGGATGAGATTCATGTGTGGAACATAGAGAGAGGTCACTCC 625
QY 151 PheGluAlaAspValThrGluLeuValAlaProGly-----GluLysPheArgLeuThr 168
Db 626 TTGGAGGTGACATCAACCAAGTGTGTCAGAGTGGGCCCTTGAACAACCTTCGGGGTCAAC 685
QY 169 IleGlyValAsnAsnGluLeuThrHisGluThrIleProProGlyLysIleThr----- 186
Db 686 ATGCGCATCAACAACACACTGACCCCTTATACCTTCCACCGGGAGCCATTTGTCTACAAAG 745
QY 187 -----ThrGlyAsnAlaThrGlyLysArgIleGlnThrTyTrpGlnHisAspPheTy 203
Db 746 ACTGATCTTCATCATGATCCCAAGGGTTACTTCGTCCAGACATAGACTTCAACTTCTTC 805
QY 204 AsnTrpAlaGlyLeuAlaArgSerIleTrpLeuTySerValProGlnGlnHisIleGln 223
Db 806 AACTATGCGGGGTGACCGGCTGTGTGTGCTGTACACACCCCTTACCACTTATTCGAT 865
QY 224 AspIleThrValValThrAspValAspGlyAspAsnGlyLeuIleAsnTrpGluValGlu 243

```

```

Db 866 GATATCACTGTACCACTGACCTGGAACCGGAGCGTTGGCTGTGAATCACTGGAATTTCT 925
QY 244 ValAlaAsnGlnThrThrGlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIle 263
Db 926 GTCAGAGGAGCAGTACCATTTCCAGCTAGAAAGTGCCTCTCTGATGAGATGCGAAATTT 985
QY 264 ValAlaAlaSerGlyValGlnGlyThrValThrIleProSerValLysLeuTrpGln 283
Db 986 GTGACCCGTGGAACAGAGGATGAGGGGTCACTTAAGTCCCGCGTCCCACTCTGGTGTG 1045
QY 284 Pro-----GluAlaAlaTyLeuTyTrpGlnLeuGlnValAsnIleValGly 298
Db 1046 CTTTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1102
QY 299 SerSerGlyAspValAlaAspThrTyTrpAsnLeuAlaThrGlyValArgThrValLysVal 318
Db 1103 ACACCTGAGTGTGTGTCTGATCTTATACCTTCCTGCGGATTCGAACAGTGGCTGTTC 1162
QY 319 AlaGlySerGlnPheLeuIleAsnGlyLysProPheTyTrpThrGlyPheGlyLysHis 338
Db 1163 ACAAGAGCAAGTTCCTCATTAATGGAAAGCTTCTTACTTCCAAAGCGCTCAACAGCAT 1222
QY 339 GluAspThrAlaValArgGlyLysGlyHisAspProAlaTyMetValHisAspPheGln 358
Db 1223 GAGGATTCAGATATCCAGAGGAGAGGCTTCGACCTGCGCTGCTGATTAAGATTTCAAC 1282
QY 359 LeuMetLysTrpIleGlyValAsnSerPheArgThrSerHisTyProTyAlaGluGlu 378
Db 1283 CTCCTCCGTGTGGCTGGGCAAAATTCCTTCCTACCAACCATACCTTACCTCGGAGAG 1342
QY 379 ValMetAspPheAlaAspArgAsnGlyIleValValIleAspGlnThrProAlaValGly 398
Db 1343 GTACTTCAAGCTCTGTGACCGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1402
QY 399 LeuAsnIleAlaLeuMetGlyValSerGluSerGlyAlaProGlnThrPheThrProAsp 418
Db 1403 ATCTGCTG-----CCCGAGAGTTT----- 1423
QY 419 AlaIleAsnAspLysThrGlnGlnLalaHisLysGlnAlaIleAspGluLeuIleAlaArg 438
Db 1424 -----GGCAAGCTGTCTCTTCGGACCACTAGAGTGAAGAGAGAGCTGTGGCGAGG 1477
QY 439 AspLysAsnHisAlaSerValValMetTrpSerIleAlaAsnGluProAlaSerHisGlu 458
Db 1478 GACAAAATAACACCTGCGCGTGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1537
QY 459 AspGlyAlaArgGluTyTrpPheGlu----- 466
Db 1538 AAACCTGCGGATATTTACTTAAGCTTATTTGTCTTCAMAGGCTGAAGACATCAGATTCC 1597
QY 467 ProLeuThrAsnLeuThrArgGluLeuAspProThrArgProIleThrPheAlaAsnVal 486
Db 1598 CCAAGAACTGAGATTACAGCTCTTGTAACTACATGCCCCGTACGTCG----- 1645
QY 487 GlyThrAlaThrTyTrpGlnLeuAspArgIleSerAspLeuPheAspValSerCysIleAsn 506
Db 1646 -----GACGTGATTTGTGTGAAC 1663
QY 507 ArgTyPheGlyTyTrpTySerGlnThrGlyAspLeuGluGluValAlaIleLeuGlu 526
Db 1664 AGTTATCTTATCTGTATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1723
QY 527 LysGluLeuHisGlyTyTrpGlnGluTyPheHisAspProIleValMetTrpGluTyGly 546
Db 1724 AGCCAGTTTGAGACTGGTATAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1783
QY 547 AlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTrpSerGluGluPheGln 566
Db 1784 GCAAGACCCGCTCTGCGGGCTTCATGAGATGCACTTCGATGATGATGATGATGATGATGATGAT 1843
QY 567 ValGlnMetLeuAspMetTyTrpHisArgValPheAsp-----ArgIleGluSerMetAla 584
Db 1844 ACAGCTCTCTCGAGAGATTATCATTTGTATCTGTGATGAGAAAGAAAGAAATGTGTATC 1903

```


Qy	585	GLYGLIHHISVALITTPAENRPhAlAAspPheGLInrPhAnLeuGLYIlelleArGyAlAsp	604
Db	1904	GGAGAGCTCACTGGAAATTTTGCTGACTTCATGACGAACCGATCAACCACTGAGAGTACA	1963
Qy	605	GLYAnLrYsLrGLYAlValPheThArGAspArGlySPoLYsAlAAlAAlHISerLeu	624
Db	1964	GGAAACAAAGAAAGGGATTCCTTCACCTGCACAGAAACCACCAAGATGGACGCTTCATTG	2023
Qy	625	ArgAlArGrTrpTrHserIle	631
Db	2024	CGAGAGAGATACTGGAGGATT	2044
RESULT 2			
LOCUS	CD014094	2095 bp	mRNA linear EST 21-OCT-2003
DEFINITION	90135027	single gene library	Homo sapiens cDNA, mRNA sequence.
ACCESSION	CD014094		
VERSION	CD014094.1	GI:37777623	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Bukacynska, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 2095)		
AUTHORS	Jin, P., Fu, G.K., Wilson, A.D., Yang, J., Chien, D., Hawkins, P.R., Au-Young, J. and Stuve, L.L.		
TITLE	PCR isolation and cloning of novel splice variant mRNAs from known drug target genes		
JOURNAL	Genomics 83 (4), 566-571 (2004)		
COMMENT	Contact: jin, P. Incyte Corporation 3160 Porter Drive, Palo Alto, CA 94304, USA Tel: 650 621 8639 Fax: 650 621 8965 Email: pjlin@incyte.com.		
FEATURES			
source	location/Qualifiers		
	1..2095		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone_lib="Single gene library"		
	/note="Vector: pDrive Cloning Vector; RT-PCR was performed using gene-specific primers flanking the open-reading frame. PCR products were subcloned into pDrive Cloning Vector and sequenced completely using M13 forward and reverse primers. Sequencing gaps were closed by re-sequencing using primers flanking the gapped areas."		
ORIGIN			
Alignment Scores:			
Pred. No.:	4,21e-129	Length:	2095
Score:	1253.50	Matches:	266
Percent Similarity:	57.51%	Conservative:	90
Best Local Similarity:	42.97%	Mismatches:	187
Query Match:	37.37%	Indels:	77
DB:	6	Gaps:	13
US-10-757-093-4 (1-634) x CD014094 (1-2095)			
Qy	42	ProGlnArGrTrHrSerSerArGluLeuValAAnLeuArpGlyLeuTrpYsPheAlaLeu	61
Db	101	CCCCAGGAGACCCCGTCGGGAGTGCAGAGAGCTGGACGGCTCTGGAGCTTC-----	154
Qy	62	AlAserGrGLYLeuAAnArPTrHr-----	154
Db	155	CGCGCCGACCTTCTGTGACACCGACGCCGGGCTTCGAGAGACGTGTACCGGCGCGG	214
Qy	76	LeuProLYsGLY-----LeuGluCySProValProAlAserTYrAsnAspIlePhe	92
Db	215	CTGTGGAGAGTACAGGCCACCGGTGCACATGCCAGTTCCTCCAGCTTCACATGCATCAGC	274
Qy	93	lIeserArGrGluIleHISArPHeIValGlyTrpValTYrGlnArGrGluValIleVal	112

[illegible]

FEATURES	source
LOCUS	CD014092
DEFINITION	1995 bp mRNA linear EST 21-OCT-2003
ACCESSION	90134967 Single gene library Homo sapiens cDNA, mRNA sequence.
VERSION	CD014092
KEYWORDS	CD014092.1 GI:37777621
SOURCE	EST.
ORGANISM	Homo sapiens (human)
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE	1 (bases 1 to 1995)
JOURNAL	Jin, P., Fu, G.-K., Wilson, A. D., Yang, J., Chien, D., Hawkins, P. R., Au-Yang, J. and Stuve, L. L.
COMMENT	PCR isolation and cloning of novel splice variant mRNAs from known drug target genes
	Genomics 83 (4), 566-571 (2004)
	Contact: Jin, P:
	Incyte Corporation
	3160 Porter Drive, Palo Alto, CA 94304, USA
	Tel: 650 621 8639
	Fax: 650 621 8965
	Email: pjim@incyte.com
	Location/Qualifiers
	1..1995
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/clone_lib="Single gene library"
	/note="Vector: pDrive Cloning Vector; RT-PCR was performed using gene-specific primers flanking the open-reading frame. PCR products were subcloned into pDrive Cloning Vector and sequenced completely using M13 forward and reverse primers. Sequencing gaps were closed by re-sequencing using primers flanking the gapped areas."

Alignment Scores:	Pred. No.:	2,62e-120	Length:	1995
Score:	1175.00	Matches:	251	
Percent Similarity:	55.92%	Conservative:	89	
Best Local Similarity:	41.28%	Mismatches:	181	
Query Match:	35.03%	Indels:	89	
DB:	6	Gaps:	12	

US-10-757-093-4 (1-634) x CD014092 (1-1995)

QY	42	ProGlnATGThrSerSerArgGluLeuValAsnLeuAspGlyLeuTrpLysPheAlaLeu	61
DB	101	CCCCAGAGAGACCCCGTGGCGGAGTGCAGAGACTGCAGCGCTCTGGAGCTTC	15
QY	62	AlaSerGlyLeuAsnAspThr-----AlaGlnProTrpThrAlaPro	75
DB	155	CGCGCGCACTTCTCTGCACACCGACCGCGGGCTTCGAGAGACGTCGACCGCGCGG	21
QY	76	LeuProLysGly-----LeuGlnCysProValProAlaSerTrpAsnAspLysPhe	92
DB	215	CTGTGGAGTACGAGCCCCACCGTGCAGATGCAGCTTCCTCCAGCTTCATATACATACGC	27
QY	93	IleSerArgGluIleHisAspHisValGlyTrpValTrpGlnArgGluValIleVal	11
DB	275	CAGAGACTGCCTGCGCGCATTTTGTTCGGCTGGCTGTGTGTGTAAGAACGGAGTACTCTG	33
QY	113	ProLysGlyTrpSerGlnGluArgTrpLeuValArgAlaGluSerAlaThrHisHisGly	13
DB	335	CCGAGCGCATGCAGCCACGAGC-----	35
QY	133	ArgIleTrpValAsnAsnAspArgLeuValAlaGlnHisValGlyGlyTrpProPheGlu	15
DB	356	-----CTGGCGCACAGAGTGGTGG-----	37
QY	153	AlaAspValThrGlnLeuValAlaProGlyGlyLysPheAspGlyLeuThrIleGlyValAsn	17
DB	374	-----CTGAGGATTTGGCGAGTGGC-----	39
QY	173	AsnGluLeuThrHisGluThrIleProProGlyLysIleThrThrGlyAsnAlaThrGly	19
DB	392	CATTC-CTATGCGCATGTGGTATCCACAG-GGTACTTT-----	42
QY	193	LysArgGlyIleGlnThrTrpGlnHisAspPheTrpAsnTrpAlaGlyLeuAlaArgSerIle	21
DB	429	-----GTCCAGAACACATATTTTGTACTTTTCAACTGACGTGCAGCTGCACCGGCTCTGTA	48
QY	213	TrpLeuTrpSerValProGlnGlnHisIleGlnAspIleThrValIleThrAspValAsp	23
DB	483	CTTCTGTGACAGACACCCACACACCTACATGATGCATACACCGTACACCAAGCGTGGAG	54
QY	233	GlyAspAsnGlyLeuIleAsnTrpGluValGluValAlaAsnGlnThrThrGlyGlnIle	25
DB	543	CAGAGACAGTGGCTGTGTGATTAACAGATCTCTGCAGAGCGGACAGAACCTGTCAAGTTG	60
QY	253	GlnIleSerValIleAspLysAspGlyAlaIleValAlaLysAsaSerGlyAlaGlnGly	27
DB	603	GAGTGCCTTTTGGATGCAGAAACAAAGTCGGTGGATGGAGTGGAGCCAGAGGC	66
QY	273	ThrValThrIleProSerValLysLeuTrpGlnPro-----GlyAlaAla	28
DB	663	CAACTTAAGTCCAGAGTGTTCAGCTCTGTGGTGGCGGTACTCGATGCAGAACGCCCTGGC	72
QY	288	TrpLeuTrpGlnLeuGlnValAsnIleValGly-----SerSerGlyAspValValAsp	30
DB	723	TATCTGATATTCATGAGGTGACGTGACATGCACACAGCGTCACTACGGGCGCTGTGTGCAC	78
QY	306	ThrTrpAsnLeuAlaThrGlyValArgThrValLysValAlaGlySerGlnPheLeuIle	32
DB	783	TCTTACACACTCCCTGTTGGGATCCGCACTGTGGCTGTGCACCAAGACCACTTCTCATC	84
QY	326	AsnGlyLysProPheTrpPheThrGlyPheGlyLysHisGluAspThrAlaValArgGly	34

Db	843	AATGGAAACCTTTTCTATTTCCACGCTGTCAACAGCATGAGATCCGACATCCGAGG	902
Qy	346	LYAGLYNHSAPPROALATYTWETWETVALHSAPRHEGLNLEMETLSELYTLEGLVALA	365
Db	903	AAGGAGCTTCGACTGGCCGCTGCTGTGAAGACTTCAACTCTCTCGCTGGCTGTGTC	962
Qy	366	ASNSERPREAAGTHSERHSISTYRPROTYRALAGLUGLVALMETASPRHEALASPRG	385
Db	963	AACGCTTTCGTAACGACCTACTACCCCTATGCGAGGAAGATGATCAGATGTGTACCGC	1022
Qy	386	ASNGLYLEVALVALHSAPGLNTHRPROALVALGLYLEUASNLALALEUWETGLY	405
Db	1023	TATGGAGTTGGTGCATGATGAGTGTCCCGCGCTGGGCTGGCCCTG-----	1078
Qy	406	VALSERGLSERGLYALAPROGLNTHRPHETHRPROASRALALEASAPLYETHRGLN	425
Db	1071	-----CCGAGTTCTTC-----ACACAGTTCTCTCG	1097
Qy	426	GLUALAHLYLGLNALILEARGLEULELALAGAPRYLASNHIALASERYAL	445
Db	1098	CATCACCACTGACAGTGTAGAMAAAGTGTGCGCTAGGAGCAAAACACCCCGCGTC	1157
Qy	446	VALMETSPSERILEALASNGLUPROLASERHSGLASPRGLVALAGLULYRPH	465
Db	1158	GTGATGTGCTGTGGCCACGAGCTGTGCTCCACTAGAAATCTGGTGGCTACTATTCG	1211
Qy	466	GLUPROLEUTHRASNLLEUTHRARGGLNLEUASPRPROTHRARPROILETHRHEALASN	485
Db	1218	AAGATGTGTATGCGCTCACACCAAACTCTTGACCCCTCCGCGCTGTGACCTT-----	1277
Qy	486	VALGLYTHRALATHRYGLNLEUASPRAGLIESERASPRLEUPHASRYALSERCYAL	505
Db	1272	GTGAGCAACTCTAATCTATGACAGACAAAGGGGCTCGGTATGTGATGTGATCTGTGTTG	1333
Qy	506	ASNPATGYTHRPHGLYTPRDPYSERGLNTHRGLYASPRLEUGLUALGLUALALALEU	525
Db	1332	AACAGCTACTACTCTTGGTATACGACACTACGGGCACTCGAATGATTCACTGCACGCTG	1391
Qy	526	GLULYVGLNLEUHSGLYTPRGLNLYSPRPHHSIARPROILEVALMETHRGLUTYR	545
Db	1392	GCCACCCAGTTTGAGAACTGTGATAGAAAGTATCAAAACCATTATTCAGACGCAAGTAT	1455
Qy	546	GLYALASPRTHYLEUHLGLYLEUHSERILEUGLYLEUPROTPRSEGLUGLUPHE	565
Db	1452	GGAGCAGAAACGATGTGACAGGGTTTCCACAGCATCACTCTGATGTCTCATGMAAGATAC	1511
Qy	566	GLNVALGIMETLEUASPRMETYRTHIARGVALPHEAP-----ARGTLEGLUSERET	583
Db	1512	CAGAAACCTGCTAGACACAGTACCATGTGGTCTGGATCAAAAACGAGAAAATACGTG	1577
Qy	584	ALAGLYGLNHSVALITPRASNPHEALASPRPHEGLNTHRASNLEUGLYILELEARGVAL	603
Db	1572	GTGGAGAGCTCATTTTGGAAATTTTCCGATTTCAAGTCACTGAACATCACACGACGAGTGTG	1633
Qy	604	ASPRGLYASNLVLYGLYVALPHETHRARGASPRGLYSPROLYSALALALAHISER	623
Db	1632	CTGGGAGATTAAGGAGGATCTTCACTCGGCGAGAGACAAACAAAGTGCAGCGTTCTT	1691
Qy	624	LEUARGALAAATGTTPRTHRSERILE	631
Db	1692	TTGCGAGAGAGATACTGGAAAGATT	1715
RESULT 4			
LOCUS	AK041058	2274 bp	mRNA linear HTCC 03-APR-2005
DEFINITION	Mus musculus adult male aorta and vein cDNA, RIKEN full-length enriched library, clone:A5507205 product:beta-glucuronidase structural, full insert sequence.		
ACCESSION	AK041058		
VERSION	AK041058.1	GI:26088231	
KEYWORDS	HTCC, CAP trapper.		
SOURCE	Mus musculus (house mouse)		

ORGANISM	Mus musculus Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1
AUTHORS	Carninci, P. and Hayashizaki, Y.
JOURNAL	High-efficiency full-length cDNA cloning
MEDLINE	Meth. Enzymol. 303, 19-44 (1999)
PUBMED	99279253
REFERENCE	10349636
AUTHORS	2
TITLE	Garcinici, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)
JOURNAL	11042159
MEDLINE	20499374
PUBMED	11042159
REFERENCE	3
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Komno, H., Akiyama, J., Nishi, K., Kitamura, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasaiwagi, K., Fujikawa, K., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Osawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000)
JOURNAL	20530913
MEDLINE	11076861
PUBMED	11076861
REFERENCE	4
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)
TITLE	5
JOURNAL	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002)
REFERENCE	6 (bases 1 to 2274)
AUTHORS	Adachi, U., Aizawa, K., Akinura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Komno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakanuma, M., Nishi, K., Nomura, K., Numasaki, R., Ono, M., Ohsato, N., Okezaki, Y., Satou, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Saeki, D., Shibata, K., Shinagawa, A., Shiraki, T., Takeda, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akehira, S., Tokabe, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
JOURNAL	Direct Submission
TITLE	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Shuto-cho, Tsukuba-shi, Yokohama, Kanagawa 230-0045, Japan [E-mail:genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216]
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL:http://genome.gsc.riken.jp/ URL:http://genome.gsc.riken.jp/. Location/Qualifiers 1..2274 /organism="Mus musculus"
FEATURES	
SOURCE	


```

Db      1625 CGGAACAGAGAGGATCTTCACTGCCAGACAGCCCAAACTCGCCTTAATTTT 1684
Qy      624  |||||
        |||||
        |||||
Db      1685 GCGAGAGAGATCTGAGGAT 1706

RESULT 5
CD014093      1853 bp      mRNA      linear      EST 21-OCT-2003
LOCUS      90135266 Single gene library Homo sapiens cDNA, mRNA sequence.
DEFINITION      CD014093
ACCESSION      CD014093
VERSION      CD014093.1 GI:37777622
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1853)
Jin, P., Fu, G.K., Wilson, A.D., Yang, J., Chien, D., Hawkins, P.R.,
Au-Yang, J. and Stuve, L.L.
PCR isolation and cloning of novel splice variant mRNAs from known
drug target genes
Genomics 83 (4), 566-571 (2004)
REFERENCE
AUTHORS      Jin, P.
CONTACT      Jin, P.
JOURNAL      Incyte Corporation
COMMENT      3160 Porter Drive, Palo Alto, CA 94304, USA
              Tel: 650 621 8639
              Fax: 650 621 8965
              Email: pj@incyte.com.
              Location/Qualifiers
                1..1853
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone_lib="Single gene library"
                  /note="Vector: pDrive Cloning Vector; RT-PCR was performed
                  using gene-specific primers flanking the open-reading
                  frame. PCR products were subcloned into pDrive Cloning
                  Vector and sequenced completely using M13 forward and
                  reverse primers. Sequencing gaps were closed by
                  re-sequencing using primers flanking the gapped areas."

ORIGIN
Alignment Scores:
Pred. No.:      8.53e-106      Length:      1853
Score:          1045.50      Matches:      227
Percent Similarity: 50.90%      Conservative: 84
Best Local Similarity: 37.15%      Mismatches: 159
Query Match:    31.17%      Indels:      141
DB:              Gaps:      13

US-10-757-093-4 (1-634) x CD014093 (1-1853)
Qy      42  ProGlnArgHsrSerArgGluLeuValAsnLeuAspGlyLeuTrpLysPheAlaLeu 61
        |||||
        |||||
        |||||
Db      101  CCGGAGAGAGCCCGGCGGAGATGCCAGAGAGCTGACGCGCTTGAGACTTC----- 154
        |||||
        |||||
        |||||
Qy      62  AlaSerGlyLeuAsnAspThr-----AlaGlnProTrpThrAlaPro 75
        |||||
        |||||
        |||||
Db      155  CGGCGGAGCTTCTCTGACCAACGCGCGGCGCTTCGAGAGCAGTGTACCGGCGCG 214
        |||||
        |||||
        |||||
Qy      76  LeuProLysGly-----LeuGluCysProValProAlaSerTyraAsnAspLysPhe 92
        |||||
        |||||
        |||||
Db      215  CTGTGGAGTCAGAGCCCGGCGGAGCATGTCAGATTCTCCAGCTTCAATGACATGAC 274
        |||||
        |||||
        |||||
Qy      93  ILeSerArgGluLeuIleAspHisValGlyTrpValTyrtYrGlnArgGluValIleVal 112
        |||||
        |||||
        |||||
Db      275  CAGAGCTGGGCTCTGCGGCACTTTTGTGCGCTGGATGGTAACAACGAGAGTATCTG 334
        |||||
        |||||
        |||||
Qy      113  ProLysGlyTrpSerGlnGlu-----ArgTyLeuValArgAlaGlnSerAlaThr 129
        |||||
        |||||
        |||||
Db      335  CCGGAGAGATGACCCAGAGACCTGCGCAACAAGAGTGTGTGAGAGATTGACATGCC--- 391

```

```

Qy      130  HisIleGlyArgIleTyraValAsnAsnArgLeuValAlaGluHisValGlyGlyTrpThr 149
        |||||
        |||||
        |||||
Db      391  -----
Qy      150  ProPheGluAlaAspValThrGluLeuValAlaProGlyGlyLysPheArgLeuThrIle 169
        |||||
        |||||
        |||||
Db      391  -----
Qy      170  GlyValAsnAsnGluLeuThrHisGluThrIleProProGlyLysIleThrThrGlyAsn 189
        |||||
        |||||
        |||||
Db      391  -----
Qy      190  AlaThrGlyLysArgIleGlnThrTyrglnHisAspPheTyraAsnTyraIleValAla 209
        |||||
        |||||
        |||||
Db      392  -----
Qy      210  ArgSerIleTrpLeuTyraSerValProGlnGlnHisIleGlnAspIleThrValAlaThr 229
        |||||
        |||||
        |||||
Db      404  -----
Qy      230  AspValAspGlyAspAsnGlyLeuIleAsnTyrgluValGluValAlaAsnGlnThrThr 249
        |||||
        |||||
        |||||
Db      410  -----
Qy      250  GlyGlnIleGlnIleSerValIleAspGluAspGlyValIleValAlaIleValAsnGly 269
        |||||
        |||||
        |||||
Db      452  TTCAGTGGAGAGTGGCTCTTTGGATGCGAACAACAAAGTGTGGAGATGGACATGGG 511
        |||||
        |||||
        |||||
Qy      270  AlaGlnGlyThrValThrIleProSerValLysLeuTrpGlnPro----- 284
        |||||
        |||||
        |||||
Db      512  ACCGAGGGCAACTTAAGGCGCAGGTGCAGCTCTGGTGGCGATCTGATGACGAA 571
        |||||
        |||||
        |||||
Qy      285  GlyAlaAlaTyrtYrLeuTyrglnLeuGlnValAsnIleValGly-----SerSerGlyAsp 302
        |||||
        |||||
        |||||
Db      572  CGCGCTGCTATCTGATTCATTTGATGAGAGTGCAGCTACTGACAGACGTCACCTGCGGCT 631
        |||||
        |||||
        |||||
Qy      303  ValValAspThrTyraAsnLeuAlaThrGlyValArgThrValLysValAlaGlySerGln 322
        |||||
        |||||
        |||||
Db      632  GTTCTGACCTTACACACTCTCTGCGGAGATGCGACCTGTGCTGACCAAGAGCCAG 651
        |||||
        |||||
        |||||
Qy      323  PheLeuIleAsnGlyLysProPheTyrtYrPheThrGlyPheGlyLysIleAspThrAla 342
        |||||
        |||||
        |||||
Db      692  TTCCTCATCATGAGAACTTCTTATTTCCAGGTCGTAACAAGCATGATGAGTGGGAC 751
        |||||
        |||||
        |||||
Qy      343  ValArgGlyLysGlyHisAspProAlaTyrtYrMetValHisAspPheGlnLeuMetLysTrp 362
        |||||
        |||||
        |||||
Db      752  ATCCGAGGGAAGGCGCTTCGACTGCGCGCTGCTGTGAAGGACTTCAACCTGCTCGTGG 811
        |||||
        |||||
        |||||
Qy      363  IleGlyAlaAsnSerPheArgThrSerHisTyrtYrProTrpAlaGluGluValMetAspPhe 382
        |||||
        |||||
        |||||
Db      812  CTGTGGTCCAAACGCTTCGATCCAGCCACTACCCCTATGACAGAGAAAGATGATGCAATG 871
        |||||
        |||||
        |||||
Qy      383  AlaAspArgAsnGlyIleValValIleAspGluThrProAlaValGlyLeuAsnIleAla 402
        |||||
        |||||
        |||||
Db      872  TGTGACCGCATGTGATGTGTGTCATGATGATGTCGCGGCGTGGCGGCGGCTG--- 928
        |||||
        |||||
        |||||
Qy      403  LeuMetGlyValSerGlnSerGlyAlaProGlnThrPheThrProAspAlaIleAsnAsp 422
        |||||
        |||||
        |||||
Db      929  -----
Qy      423  LysThrGlnGluAlaHisLysGlnAlaIleArgGluLeuIleAlaArgAspLysValAsnHis 442
        |||||
        |||||
        |||||
Db      947  GTTCTCTGTCATACCAATGAGTGAAGAGAGTGTGTGAGGACCAAGAACAC 1006
        |||||
        |||||
        |||||
Qy      443  AlaSerValValMetTrpSerIleAlaAsnGluProAlaSerHisGluLysAspGlyAlaArg 462
        |||||
        |||||
        |||||
Db      1007  CCGCGGTGTATATGTGTGTGTGTGCGCAACAGCCGCGTCCCACTAAGATGTGCGG 1066
        |||||
        |||||
        |||||
Qy      463  GluTyrtPheGluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThr 482
        |||||
        |||||
        |||||
Db      1067  TACTACTTGAAGATGTGATGTGCTGCTACACCAAAATCTTGACCCCTCCGCGCTGTGACC 1126

```

Qy 483 PheAlaAsnValGlyThrAlaThrThyGlnLeuAspArgLleSerAspLeuPheAspVal 502
 Db 1127 TTT-----GTAGCACTCTAATGATGACAGACAGAGGGGGCTCCGTATGTGATGTG 1180
 Qy 503 SerGylleAsnArgThyPheGlyTrpTySerGlnThrGlyAspLeuGluAlaGlu 522
 Db 1181 ATCTTTGAAACAGTACTACTCTGTATCAGCACTACGGGACCTGGAGTGTGATTCAG 1240
 Qy 523 AlaAlaLeuGluLeuHisGlyTrpGlnGluValPheHisArgProLleValMet 542
 Db 1241 CTGACGCTGCCACCCAGCTTTGAGAACTGTATTAAGATACAGAACCCCATTAATCAG 1300
 Qy 543 ThrGlyTrpGlyAlaAspThrLeuAlaGlyLeuHisSerLleLeuGlyLeuProTrpSer 562
 Db 1301 AGCGAGTATGGAGCAGAAACGATTCAGGGTTTCCAGCATCCACCTCTATATTCTACT 1360
 Qy 563 GluGluPheGlnValGlnMetLeuAspMetTyHisArgValPheAsp-----ArgLle 580
 Db 1361 GAAGAGTACAGAAAAGTCTGTAGACAGTACCATCTGGGCTGTGATCAAAAACGAGA 1420
 Qy 581 GluSerMetAlaGlyGluHisValTrpAsnPheAlaAspPheGlnTrpAsnLeuGlyLle 600
 Db 1421 AAATACGTGTGGAGAGACTCATTTGGATTTTGCCTTTCATGACTGAAACAGTACCG 1480
 Qy 601 IleArgValAspGlyAsnLysGlyValPheThrArgAspArgLysProLysAlaAla 620
 Db 1481 ACGAGAGTCTGGGGATTAAGAGGATCTTCACTCGGAGAGACAAACCAAAAGTGCA 1540
 Qy 621 AlaHisSerLeuArgAlaArgTrpThrSerLle 631
 Db 1541 GCGTTCCTTTGCGAGAGATCTCGAAGAT 1573

RESULT 6
 CR593823
 LOCUS full-length cDNA clone CS0DL001YM21 of B cells (Ramos cell line)
 DEFINITION Cot 25-normalized of Homo sapiens (human).
 ACCESSION CR593823
 VERSION CR593823.1 GI:50474630
 KEYWORDS HTC; CNSLT cDNA.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euteheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 AUTHORS Li, W.-B., Gruber, C., Jessee, J., and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished
 CONTACT: Feng Liang Email: fliang@life.techn.com URL: http://fulllength.invitrogen.com/InvitrogenCorporation1600
 REMARK

REFERENCE
 AUTHORS Paraday Avenue
 TITLE Genoscope.
 JOURNAL Direct Submission
 SUBMITTED (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr
 COMMENT - Web : www.genoscope.cns.fr)
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
 was normalized. Library was constructed by Life Technologies, a
 division of Invitrogen.

FEATURES
 source location/Qualifiers
 1..1377

ORIGIN
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DL001YM21"
 /isue_type="B cells (Ramos cell line) Cot 25-normalized"
 /plasmid="pCMVSPORT_6"

* Alignment Scores:

	Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	DB:
US-10-757-093-4 (1-634) x CR593823 (1-1377)	8.28e-95	946.00	61.48%	45.93%	28.21%	3
						Gaps: 6
Qy 236 GlyLeuIleAsnTyrgluValAlaAsnGlnThrGlyGlnIleGlnIleSer 255						
Db 1 GGGGTGTGATTAACGAACTCTGTCAAGGGGAGAACCTGTCAAGTGAAGGCT 60						
Qy 256 ValIleAspGluAspGlyAlaIleValAlaLysAlaSerGlyAlaGlnGlyThrValThr 275						
Db 61 CTTTGGATGCAAAAACAAGTCGTGGGAATGGAGTGGGACCCAGGCCAAGCTTAAG 120						
Qy 276 IleProSerValLysLeuTrpGlnPro-----GlyAlaAlaLysLeuTrp 290						
Db 121 GTCCAGGTGTGAGCTCTGTGGGCGCTACCTGATGCAGAACGCCCTGTCTATCTAT 180						
Qy 291 GlnLeuGlnValAsnLleValGly-----SerSerGlyAspValValAspThrTyra 308						
Db 181 TCATTGAGAGTGCAGCTGACTGCAGACAGTCACTGGGGCTGTGTCTGATCAACA 240						
Qy 309 LeuAlaThrGlyValArgThrValLysValAlaGlySerGlnPheLeuIleAsnGlyLys 328						
Db 241 CTCCCTGTGGGAGATCGCAGCTGTGTGTGCACCAAGCCATTTCTCATCAATGGGAAA 300						
Qy 329 ProPheTyPheThrGlyPheGlyLysHisGluAspThrAlaValArgGlyLysGlyHis 348						
Db 301 CTTTCTATTTCACAGGTGTCAACAGCATGAGATGGGAGATCCGAGGAAAGGCTTC 360						
Qy 349 AspProAlaTyMetValHisAspPheGlnLeuMetLysTrpLleGlyAlaAsnSerPhe 368						
Db 361 GACTGCCCTCTGTGTGAAGGACTTCAACCTGCTTCCGTGGTGGCAACGCTTTC 420						
Qy 369 ArgThrSerHisTyProTyArgLysGluValMetAspPheAlaAspArgAsnGlyLle 388						
Db 421 CGTACAGCCACTTACCCCTATGACAGAGAGAGAGTGCAGATGTGACCGCTATGGGATT 480						
Qy 389 ValValIleAspGluThrProAlaValGlyLeuAsnIleAlaLeuMetGlyValSerGlu 408						
Db 481 GTGTGATGATGATGATGTCGCCGCGTGGCGCTGGCGCTG----- 519						
Qy 409 SerGlyAlaProGlnThrPheThrProAspAlaIleAsnAspLysThrGlnGluAlaHis 428						
Db 520 -----CGCACTTCTTC-----AACCAAGTTTCTTCATCACAC 555						
Qy 429 LysGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHisAlaSerValIleMetTrp 448						
Db 556 ATGCAAGTATGAGAAAGATGTGTGTCAGGACAAAGCAACCCCGCGTGTATGTGG 615						
Qy 449 SerIleAlaAsnGluProAlaSerHisGluAspGlyAlaArgGlyLysPheGluProLeu 468						
Db 616 TCTGTGGCAACGAGCCCTGCTCCACCTAGAACTGTGCTACTTGAATGATGGT 675						
Qy 469 ThrAsnLeuThrArgGlnLeuAspProThrArgProLleThrPheAlaAsnValGlyThr 488						
Db 676 ATCGTCAACCAAACTTGAACCCCTCCGCGCTGTGACTTT-----GTAGCAAC 729						
Qy 489 AlaThrTyrgluLeuAspArgLleSerAspLeuPheAspValSerGylleAsnArgTy 508						
Db 720 TCTAATATGACAGACAGAAAGGGGCTCCGATGTGATGTGATCTTTGAACAGCTAC 789						
Qy 509 PheGlyTrpTySerGlnThrGlyAspLeuGluGluAlaGluAlaLeuGlyLysGlu 528						
Db 790 TACTCTTGATGATCAAGTACAGGCGACCTGAGTGTGATGACGTGCGACCCACAG 849						
Qy 529 LeuHisGlyTrpGlnGluLysPheHisArgProLleValMetThrGlyTrpGlyAlaAsp 548						
Db 850 TTTGAACTGTGATTAAGAAAGTATCAAGGCCCATTTATCAAGCGAGTATGGACGAA 909						

QY 549 ThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTyrSerGluGluPheGlnValGln 568
 Db 910 ACATTGCGAGGCTTTCACGACGATCCACCTCTATGCTCTGAAAGTCCAGAAAGT 969
 QY 569 MetLeuAspMetTyrHisArgValPheAsp-----ArgIleGluSerMetAlaGlyGlu 586
 Db 970 CTCCTAGACAGACATCTGCGTCTGATCAAAAACGAGAAATACGTGTTGGAGAG 1029
 QY 587 HisValTyrPhePheAlaAspPheGlnThrAsnLeuGlyIleLeuArgValAspGlyAsn 606
 Db 1030 CTCATTGGAAATTTGGCATTTTCATGACGAAACAGTCCAGCAGAGAGTCTGGGGAAT 1089
 QY 607 LysAlaGlyValPheThrArgAspArgGlyProLysAlaIleAlaHisSerLeuArgAla 626
 Db 1090 AAAAAGGAGATCTTCTACTCGGACAGACACAAAAGTCCAGCTTCTTTGCCAGAG 1149
 QY 627 ArgTyrThrSerIle 631
 Db 1150 AGTACTGGAAGATT 1164
 RESULT 7
 CDS03076 1124 bp mRNA linear EST 12-JUN-2003
 LOCUS CDA60-C07.xid-c SHGC-CDA Gasterosteus aculeatus cDNA clone
 DEFINITION CDA60-C07 5', mRNA sequence.
 ACCESSION CDS03076
 VERSION CDS03076.1 GI:31432977
 KEYWORDS EST.
 SOURCE Gasterosteus aculeatus (three spined stickleback)
 ORGANISM Gasterosteus aculeatus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
 Gasterosteidae; Gasterosteus.
 1 (bases 1 to 1124)
 Kingsley, D.M., Peichel, C., Balabhadra, S., Grimwood, J., Dickson, M.,
 Schmutz, J., and Myers, R.M.
 Expressed sequence tags from Gasterosteus aculeatus
 Unpublished (2003)
 Contact: Kingsley, DM
 HMMI and Department of Developmental Biology
 Stanford University School of Medicine
 Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA
 Tel: 650 725 5954
 Fax: 650 725 7739
 Email: kingsley@cmgm.stanford.edu
 Plate: 60
 High quality sequence stop: 782.
 Location/Qualifiers
 1..1124
 /organism="Gasterosteus aculeatus"
 /mol_type="mRNA"
 /strain="Salinas river, CA"
 /db_xref="taxon:69293"
 /clone="CDA60-C07"
 /sex="mixed male and female"
 /tissue_type="heads and internal organs combined"
 /dev_stage="adult"
 /note="Vector: lambda ZAP Express/pBK-CMV; Site 1: EcoRI
 (5' adaptor); Site 2: XhoI (3' linker primer); The mixed
 organ cDNA library was generated using the ZAP-cDNA method
 by Stratagene. First strand cDNA synthesis was primed with
 a 50 bp linker primer containing an oligo dt sequence
 preceded by a synthetic XhoI site. 5 prime adaptors were
 used containing an EcoRI cohesive end. The finished cDNAs
 were inserted in to the ZAP express vector
 unidirectionally in the sense orientation with respect to
 the lacZ promoter of pBK-CMV. An amplified library was
 prepared from approximately 3 million primary clones in
 the lambda ZAP Express vector. In vivo excision was then
 used to generate individual pBK-CMV phagemid clones for
 EST sequencing."

ORIGIN
 Alignment Scores:
 Pred. No.: 7,26e-79 Length: 1124
 Score: 803.00 Matches: 164
 Percent Similarity: 60.748 Conservative: 48
 Best Local Similarity: 46.994 Mismatches: 115
 Query Match: 23.944 Indels: 23
 DB: 6 Gaps: 5
 US-10-757-093-4 (1-634) x CDS03076 (1-1124)
 QY 284 ProGlyAlaAlaTyrLeuTyrGlnLeuGlnValAsnIleValGlySerSerGlyAspVal 303
 Db 8 CCAACCCAGGTTACTCTTACTCTTTGAGAGTTGCTTACTGCTGCTAGAGATCTGCT 67
 QY 304 Val--AspThrTyrAsnLeuAlaThrGlyValaArgThrValaAlaGlySerGln 322
 Db 68 TCCACGACGCTGTAACCTACACGATCGGCATCCGACGGTCGACGCTTACACACCCAG 127
 QY 323 PheLeuIleAsnGlyLysProPheTyrPheThrGlyPheGlyLysHisGluAspThrAla 342
 Db 128 TTCCTCATCAAAACAAAGCCCTTCTACTCCACGAGTAATTAACAACAGAGCTGTGAT 187
 QY 343 ValArgGlyLysGlyHisAspProAlaTyrMetValHisAspPheGlnLeuMetLysTyr 362
 Db 188 ATTGAGGCAAGGCTTGACCTGCCCCCTCATGCTGTAAGACTTAACTTATGAAGTGG 247
 QY 363 IleGlyAlaAsnSerPheArgThrSerHisTyrProTyrAlaGluGluValMetAspPhe 382
 Db 248 TTGGGGGCCCAACTCGTTCCGACACCACTACCTTATGACAGAGAGATCTGCAGATG 307
 QY 383 AlaAspArgAsnGlyIleValaIleAspGluThrProAlaValaGlyLeu--AsnIle 401
 Db 308 TGTGACCGCATGCGATCGTGTGATGACGATGCCGGCGCTGGCATTAACCAAT 367
 QY 402 AlaLeuMetGlyValSerGluSerGlyAlaProGlnThrPheThrProAspAlaIleAsn 421
 Db 368 CGCATTTTGA----- 379
 QY 422 AspLysThrGlnGluAlaHisLysGlnAlaIleArgGluLeuIleAlaArgAspLysAsn 441
 Db 380 AACGCTCCTTAACCATCATCATCGTGTGATGACGAGCTGTGATCGTGGCAAGAAC 429
 QY 442 HisAlaSerValaValMetTyrPheIleAlaAsnGluProAlaSerHisGluAspGlyAla 461
 Db 440 CATCCCTGTGTGATGATGTGTGATGAGCCCAAGACCGGCTGCAGAGATGCTCTGCT 499
 QY 462 ArgGluTyrPheGluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIle 481
 Db 500 GATTACTATTTCAAACCTTGATTAACATACCAAGAAATTGATTCACCCGCCCTGTC 559
 QY 482 ThrPheAlaAsnValaGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPheAsp 501
 Db 560 ACTTTT-----ATCACAGACAGTACTATGACAGGATTAAGAGGCTCCTACGTGAC 613
 QY 502 ValSerCysIleAsnArgTyrPheGlyTyrTyrSerGlnThrGlyAspLeuGluGluAla 521
 Db 614 GTCATCTGGCTAAACGTTACTCTCTGTTACATGACCCGGGCCACCGCGAGTCAATC 673
 QY 522 GluAlaAlaLeuGluLysGluLeuHisGlyTyrGlnGluLysPheHisArgProIleVal 541
 Db 674 CCCATCCAGTCAACACTCGTTTGAGAACTGTATCGAAGAAAGTACCAAAACCATCATC 723
 QY 542 MetThrGluTyrGlyAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTyr 561
 Db 734 CAAGAGGAATACGAGCGAGTGTGCGCGGCTTCAACGTGATCCACCGGATGTTT 793
 QY 562 SerGluGluPheGlnValGlnMetLeuAspMetTyrHisArgValaPheAsp-----Arg 579
 Db 794 ACTGAGAGATACCGAAGTATGATCGCAGAGCTACCAACCTGTGTGACCGAAGAGG 853
 QY 580 IleGluSerMetLysGlyGluHisValTyrAsnPheAlaAspPheGlnThrAsnLeuGly 599

```

Db      854  AAGCAGTACGTCATTCGCGCAACTCATCTGAACTTTGACAGCTTCATGACCCACAGGAG 913
Qy      600  ILeIleArgValaAspGlyAsnLysGlyValaPheThrArgAspArgLysProLysala 619
Db      914  ATCATCGCGGTGTGGGGAACAAAAGGCTGTCTCTACAGCAGGCA-AGGAGCCCAAGCAG 972
Qy      620  AAlaAlHisSerLeuArgAlaArgTTP 628
Db      973  CAGCATCATCTGAAGGAGAGTACTGG 999

RESULT 8
BX63460
LOCUS   BX63460 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
DEFINITION Homo sapiens cDNA clone CS0DL001YM21 5-PRIME, mRNA sequence.
ACCESSION BX63460
VERSION  BX63460.2 GI:46291531
KEYWORDS EST.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1051)
          Li, W.-B., Gruber, C., Jessee, J. and Polyes, D.
          Full-length cDNA libraries and normalization
          Unpublished (2001)
          On May 5, 2003 this sequence version replaced gi:30372720.
COMMENT   Contact: Genoscope
          Genoscope - Centre National de Sequencage
          2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
          Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
          1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
          end enriched, double-strand cDNA was digested with Not I and cloned
          into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
          was normalized. Library was constructed by Life Technologies, a
          division of Invitrogen. This sequence belongs to sequence cluster
          4169.r
          For more information about this cluster, see
          http://www.genoscope.cns.fr/cdna?cs=CS0DL001YM21&c=4169.r.
FEATURES
          source
          location/Qualifiers
            1..1051
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="CS0DL001YM21"
            /cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
            /cell_line="RAMOS CELL LINE"
            /clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT
            25-NORMALIZED"
            /note="1st strand cDNA was primed with a NotI-oligo(dT)
            primer. Five prime end enriched, double-strand cDNA was
            digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
Alignment Scores:
Pred. No.: 2.72e-78 Length: 1051
Score: 797.50 Matches: 159
Percent Similarity: 59.50% Conservative: 57
Best Local Similarity: 43.80% Mismatches: 121
Query Match: 23.78% Indels: 26
DB: 5 Gaps: 6
US-10-757-093-4 (1-634) x BX63460 (1-1051)
Qy      236  GlyLeuIleAsnTyrGluValaGluValaAlaAsnGlnThrThrGlyGlnIleGlnIleSer 255
Db      1 GGGCTGGGATTCACATCTCTCTCAAGGACGTAACCTGTTCAAGTTGGAAGTGCT 60
Qy      256  ValIleAspGluAspGlyAlaIleValaAlaLysAlaSerGlyAlaGlnGlyThrValaThr 275
Db      61 CTTTGGATGCAAAAACAAAGTCTGTGGCGAATGGAGCTGGAGCCAGGCGCAACTTAAG 120

```

```

Qy      276  IleProSerValLysLeuTrpGlnPro-----GlyAlaAlaTyrLeuTyr 290
Db      121  GGGCAGGATGTCAGGCTTGTGGTGGCCGTAAGTACAGCAAGCCCTGCTTATCTGAT 180
Qy      291  GlnLeuGlnValaAsnIleValaGly-----SerGlyAspValaValaAspThrTyrAsn 308
Db      181  TCATTGGAGGTGACAGCTACGACAGACAGACGTCAGGGGCTGTGATCTTACATCTACACA 240
Qy      309  LeuAlaThrGlyValaArgThrValaLysValaAlaGlySerGlnPheLeuIleAsnGlyLys 328
Db      241  CTCCTGTGGGATCCGACCTGTGCTGTACCAAGCCAGCTTCTCATATGAGGAAA 300
Qy      329  ProPheTyrPheThrGlyPheGlyLysHisGlyLeuPheThrAlaValaArgGlyLysGlyHis 348
Db      301  CTTTCTCTATTTCCACGCGGTGTCAACAGCATAGATGGCAGACATCCAGGAAAGGCTTC 360
Qy      349  AspProAlaTyrMetValaHisAspPheGlnLeuMetLysTrpIleGlyAlaAsnSerPhe 368
Db      361  GACTGGCGCGTGTGTGAAGACTTCAACTGCTTCGCTGGCTGTGGCCACAGCTTTC 420
Qy      369  ArgThrSerHisTyrProTyrAlaGlnGluValaMetAspPheAlaAspArgAsnGlyIle 388
Db      421  CGTACAGGCACTACCTCCATGACAGAGATGATGATGTGATGACCGCTATGGGATT 480
Qy      389  ValValIleAspGluThrProAlaValaGlyLeuAsnIleAlaLeuMetGlyValaSerGlu 408
Db      481  GTGTCTATGATGATGATGTGCCGCGCTGGCCCTGGCGCTG----- 519
Qy      409  SerGlyAlaProGlnThrPheThrProAspAlaIleAsnAspLysThrGlnGluAlaHis 428
Db      520  -----CCGCAAGTTCTTC-----AACACAGTTTCTGTGATCACCAC 555
Qy      429  LysGlnAlaIleArgGlnLeuIleAlaArgAspLysAsnHisAlaSerValaValaMetTrp 448
Db      556  ATGCAGGTATGAAAGATGGTGGTGGAGACAAAGAACACCCCGCGTCTGATGTGG 615
Qy      449  SerIleAlaAsnGluProAlaSerHisGlyAspGlyAlaArgGluTyrPheGlnProLeu 468
Db      616  TCTGTGGCCACAGAGCTGCTGCCACCTAGAAATGCTGCTACTTGAAGATGGTG 675
Qy      469  ThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnValaGlyThr 488
Db      676  ATCGCTCACACCAAAATCCTTGAGACCCCTCCGCGCTGTGACTTT-----GTGAGCAAC 729
Qy      489  AlaThrTyrGlnLeuAspArgLysSerAspLeuPheAspValaSerCysIleAsnArgTyr 508
Db      730  TCTAACTATGACAGCAGACAAAGGGGCTCGTATGTGATGTATCTGTTGAACAGCTAC 789
Qy      509  PheGlyTyrTyrSerGlnThrGlyAspLeuGlnGluAlaGluAlaLeuGlnLysGlu 528
Db      790  TACTCTGGATACAGACATACGAGGACCTGAGATGATTCACTGACGCTGCCACCCAG 849
Qy      529  LeuHisGlyTyrGlnGlnLysPheHisArgProIleValaMetThrGlnLysGlyAlaAsp 548
Db      850  TTTGAGAACTGTATATGAAGATATCAGAGCCCATTTATTCAGAGCAGATATGAGACAGA 909
Qy      549  ThrLeu-AlaGlyLeuHisSerIleLeuGlyLeuProTyrSerGlnGluPheGlnValaGlu 568
Db      910  ACGATTTGGAGGTTTCCACCAAGATCCACTGTGATGTCTACAGAAAGATACAGAAAR 969
Qy      568  MetLeuAspMetTyrHisArgValaPheAspArgIleGlnSerMetAlaGlyGluHis-- 587
Db      970  TTYTCTAAGACAGTACATCTGGGTGTGATAAAMAGMAAAATATGTGTGRRRCT 1029
Qy      588  -ValTTP 589
Db      1030  CATTTGG 1036

RESULT 9
CL486845/c
LOCUS   CL486845 921 bp DNA linear GSS 01-APR-2004
DEFINITION SAIL_443_B06.v2 SAIL Collection Arabidopsis thaliana genomic clone

```

FEATURES	source	Location/Qualifiers
		1..921
		/organism="Arabidopsis thaliana"
		/mol_type="genomic DNA"
		/ecotype="Columbia"
		/db_xref="taxon:3702"
		/clone="SAIL 443 B06.v2"
		/clone_1bp="SAIL Collection"
		/note="Tr-DNA left border sequences were isolated using a modified Tail-PCR strategy"

D6		493	TGTAAATGTTTGGCAAGCCTCACACCCGATTACCATTCAGCGATCTCTTGATGTGCCTGCTCCT	433
OY		505	EAAATAAGTGTPhBgG1TTrPytSeegInPnrgUyAarBueGluJuaIaaIale	525
D6		433	GAAACCGTATTATGAATGATGTATGTATGCCAACGGCGGATTTGGAAAACGGCAGAAGACTACT	374
OY		525	UGUUYUGLyeHniEgUlYTrgIngUuyvPhneIAxGProIIeValmetrIgUyT	545
D6		373	GGAAAAAGAACTTCTGGCGCTGGCGAGAGAAACSTGCATCAGCCGATTATCATCACCAGATA	314
OY		545	rgUyAlaarPThryLeuIaGlyLeuHisserrlleUglyLeuProTpbergIngUyph	565..
D6		313	CGGCGTGGATACGTTAGCCGGCGCTGCACATCAATGTAACCGACAATGTGGAGTGAAGATA	254
OY		565	eGlnValGImetleuAarVetyrHibaGvalPheAarPgTllectusemectAlaGl	585
D6		253	TCAGTGTGCATCGGCTGGATGTGTATCAACCGCGTCTTGATGCGCTGACGCGCGTGGCG	194
OY		585	yGUnHiseValTPraasPhelaAarPheGIntPrrAenUeGlyTelletlaayValargI	605
D6		193	TGAACAGATATGAATTTCCGCAATTTTGGACCTCCGAAGCAATATGCGCGTGGCGCG	134
OY		605	YAanUySLyeG1yValPherThiraArPaArgLySProLyASalaIahIsserleuAr	625
D6		133	TAAACAGAAAGGATGCTTCACTCCGACCCGCAACCAAGTCGCGCTTTTAATGAAG	74
OY		625	gaIa	626
D6		73	AAGC	70
RESULT_10				
CD503098				
LOCUS				
DEFINITION			CD460-D07.xid-c SHGC-CDA Gasterosteus aculeatus cDNA clone	
ACCESSION			CD460-D07 5', mRNA sequence.	
VERSION			CD503098	
KEYWORDS			CD503098.1 GI:31433163	
ORGANISM			EST.	
SOURCE			Gasterosteus aculeatus (three spined stickleback)	
COMMENT			Gasterosteus aculeatus Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes; Gasterosteidae; Gasterosteus. 1 (bases 1 to 1138) Kingsley,D.M., Pelchei,C., Balabhadra,S., Grilwood,J., Dickson,M., Schmutz,J. and Myers,R.M. Submitted (2003) Expressed sequence tags from Gasterosteus aculeatus Unpublished (2003) Contact: Kingsley, DM HHMI and Department of Developmental Biology Stanford University School of Medicine Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA Tel.: 650 725 5954 Fax: 650 725 7739 Email: kingsley@cmgm.stanford.edu Plate: 60 High quality sequence stop: 782. Location/Qualifiers 1..1138 /organism="Gasterosteus aculeatus" /mol_type="mRNA" /strain="Salinas river, CA" /db_xref="taxon:69293" /clone="CD460-D07" /sex="mixed male and female" /tissue_type="heads and internal organs combined" /dev_stage="adult" /notes="Vector: lambda ZAP Express/pBK-CMV; Site 1: EcoRI (5' adaptor); Site 2: XhoI (3' linker primer); The mixed organ cDNA library was generated using the ZAP-cDNA method by Stratagene. First strand cDNA synthesis was primed with	

a 50 bp linker primer containing an oligo dt sequence preceded by a synthetic XhoI site. 5 prime adaptors were used containing an EcoRI cohesive end. The finished cDNAs were inserted in to the ZAP express vector unidirectionally in the sense orientation with respect to the lacZ promoter of pBK-CMV. An amplified library was prepared from approximately 3 million primary clones in the lambda ZAP Express vector. In vivo excision was then used to generate individual pBK-CMV phagemid clones for EST sequencing."

ORIGIN

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	770.50	166	47	111	25	5
Percent Similarity:	61.56%					
Best Local Similarity:	47.98%					
Query Match:	22.97%					

US-10-757-093-4 (1-634) x CD503098 (1-1138)

```

QY 284 ProGlyAlaAlaTyrLeuTyrGlnLeuGlnValAsnIleValGlySerSerGlyAspVal 303
Db 8 CCGAACCCAGGTTACCTTACTCTTGTGAGGTGTCTTACTGCTGTGAGGATCTGCT 67
QY 304 Val---AspThrTyrAsnLeuAlaThrGlyValArgThrValAlaGlySerGln 322
Db 68 TCACAGACGTCGTACCTTACCAAGTCGACGCGACGCTTACAGAGATCTGACAGC 127
QY 323 PheLeuIleAsnGlyLysProPheTyrPheThrGlyPheGlyLysIleAspThrAla 342
Db 128 TTCTCTCAACAAAGAGCCCTTCTTACTTCCAGAGTAATATAACGAGGACTCTGAT 187
QY 343 ValArgGlyLysGlyHisAspProAlaTyrMetValHisAspPheGlnMetLysTrp 362
Db 188 ATTGAGGCAAGAGCTGAGCTGCGCCCTCATGCTGAAGACTTTAACTTATGAAGTGG 247
QY 363 IleGlyAlaAsnSerPheArgThrSerHisTyrProTyrAlaGlnGluValMetAspPhe 382
Db 248 TTGGGGGCGCAACTGTTCCGACACAGCACTTATGACAGAGATCTGCGAGATG 307
QY 383 AlaAspArgAsnGlyIleValValIleAspGlnThrProAlaValGlyLeu---AsnIle 401
Db 308 TGTGACCGCGCATGCTGCTGTGATAGACGAGTGCCTGGGCAATAAAGACATT 367
QY 402 AlaLeuMetGlyValSerGlySerGlyAlaProGlnThrPheThrProAspAlaIleAsn 421
Db 368 CGCAGTTTGGG----- 379
QY 422 AspLysThrGlnGlnIleHisIleGlnAlaIleArgGlnLeuIleAlaArgAspLysAsn 441
Db 380 AAGCGCTCTTAACCATCAGTCGTGTCATGAGAGAGTGGTAGTGGGAGCAAGAAC 439
QY 442 HisAlaSerValValMetTrpSerIleAlaAsnGlnProAlaSerHisGlnLysGlyAla 461
Db 440 CATTCCCTGTGTGTCATGTGTGTCAGTGGCAATAGACCGGCTGCAGAGATGCTCTGCT 499
QY 462 ArgGlyTyrPheGlnLysProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIle 481
Db 500 GATTACTATTTCAAACCTTGATMAACATACCAAGAAATGGATGCAACCGGCGCGTC 559
QY 482 ThrPheAlaAsnValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPheAsp 501
Db 560 ACTTTT-----ATCACAGACAGTAACTATGCCAGGATTAAGGGGCTCCCTACGTGAGC 613
QY 502 ValSerCysIleAsnArgTyrPheGlyTyrPyrTyrSerGlnThrGlyAspLeuGlnGluAla 521
Db 614 GTTATCTGTGTAACGTTACTTCTCTGTATCCATGACCCGCGACCCGAGGTGATC 673
QY 522 GluAlaAlaLeuGlnLysGlnLeuHisGlyTyrPheGlnGlnLysPheHisArgProIleVal 541
Db 674 CCCATCCAGCTCAACACTGATTGAGAACTGTGACGAGAAAGTACCAAGAACCCATCATC 733

```

```

QY 542 MetThrGlnTyrGlyValAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTrp 561
Db 734 CAGAGGAAATACGAGCGGATGGCGGCGGCTTCAACAGATCCACCCGATGTT 793
QY 562 SerGlnGlnPheGlnValGlnMetLeuAspMetTyrHisArgValPheAsp-----Arg 579
Db 794 ACTGAGAGATACCAAGATTAAGTCTGTCAGAGACTTACCAAGAGTGTTCACAGAAAGA 853
QY 580 IleGlnSerMetAlaGlyGlnHisValTyrAsnPheAlaAspPheGlnThrAsnLeuGly 599
Db 854 AG-CAGTACGTATCGCGCAACTCTGAACTTTCAGACTTCATGACACACAGAGG 912
QY 600 IleIleArgValAspGlyAsnLysGlyValAlaPheThrArgAspArgLysProLysAla 619
Db 913 ATCATCGGTGTGTGGGGAACAGAG-GGTGTCTTACAGGAGAAAGCGCAAGCA- 970
QY 620 AlaAlaHisSerLeuArg 625
Db 971 GCAGCATTCATCTTAAGG 988

```

RESULT 11

BX745933/c

LOCUS BX745933 857 bp mRNA linear EST 18-NOV-2003

DEFINITION

BX745933 XGC-gastrula Xenopus tropicalis cDNA clone Tga066d19 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 857
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="Tga066d19"
/dev_stage="gastrula (stages 10.5-12 mixed)"
/lab_host="Escherichia coli XL1-blue"
/clone_id="XGC-gastrula"
/note="Vector: PCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dt primed from 5' end of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into PCS107 with EcoRI at the 5' end and NotI at the 3' end."
Location/Qualifiers

This sequence is from a Xenopus Gene Collection (XGC) library constructed by Aaron M. Zorn.
cDNA was oligo dt primed from 5' end of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into PCS107 with EcoRI at the 5' end and NotI at the 3' end.
Vector: PCS107; Site 1: EcoRI; Site 2: NotI
Host: Escherichia coli XL1-blue.
Location/Qualifiers

ORIGIN

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:
Score:	1.9e-70	857	144	44
Percent Similarity:	726.50			
Best Local Similarity:	63.95%			
	48.98%			

QY 255 SerValIleAspGluAspGlyAlaIleValAlaIleValAlaSerGlyAlaGlnGlyThrVal 274
 DB 466 GAACCTCGGATCCGATCAACAGCGTGTGCAACGACCAAGCCAGCGGACTTGG 525
 QY 275 ThrIleProSerValIleLeuTrpGlnProGlyAlaIleValIleLeuTrpGlnLeuGlnVal 294
 DB 526 CAAGTGTGATATCCGACCTCTGGCAATCGGTGAAGTTATCTATCACTGATGCGTC 585
 QY 295 AsnIleValIleGlySerSerGlyAspValValAspThrTyrAsnLeuAlaThrGlyValArg 314
 DB 586 -----ACGCCAAAGCCAGACAGAGTGTGATATCAACCGCGCGCGCATCCGG 639
 QY 315 ThrValIleValAlaGlySerGlnPheLeuIleAsnGlyValAspProPheTyrPheThrGly 334
 DB 640 TCAGTGGCAGATGAAGGGCGAAGCAGTTCCTGATCAACCAACCGTTCTTACTTACTGCG 699
 QY 335 PheGlyValIleGlnAspThrAlaValArgGlyValGlyValIleAspProAlaTyrMetVal 354
 DB 700 TTGGCGCGTCAATGAGATGGCGATTTGGCGGCAAGATTGATTAACGTCTGATGATG 759
 QY 355 HisAspPheGlnLeuMetIleTyrIleGlyAlaAsnSerPheArgThrSer 371
 DB 760 CACGATACCGCATTTATGACTGATTGGGCCCACTCTCACTCACTGCTG 810

RESULT 13 BQ941196 906 bp mRNA linear EST 21-AUG-2002
 LOCUS BQ941196
 DEFINITION AGENCOURT_8741562 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6420519
 5', mRNA sequence.

ACCESSION BQ941196
 VERSION BQ941196.1 GI:22356674
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 906)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgsrbs-remail.nih.gov
 Tissue Procurement: DCTD/DPF/Gazdar
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
 http://image.llnl.gov
 Plate: LUCM2594 row: 9 column: 16
 High quality sequence stop: 765.

FEATURES

source

1..906
 location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6420519"
 /tissue_type="large cell carcinoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 18"
 /note="Organ: lung; Vector: pOTB7; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGCACGAG(G). Library constructed by Ling Hong in the
 laboratory of Gerald M. Rubin (University of California,
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
 Superscript II RT (Life Technologies). Note: this is a
 NIH_MGC Library."

ORIGIN
 Alignment Scores:
 Pred. No.:

4.77e-68 Length: 906

Score: 705.50 Matches: 142
 Percent Similarity: 61.44% Conservative: 46
 Best Local Similarity: 46.41% Mismatches: 98
 Query Match: 21.03% Indels: 20
 DB: 5 Gaps: 4

US-10-757-093-4 (1-634) x BQ941196 (1-906)

QY 331 TyrPheThrGlyPheGlyValIleGlnAspThrAlaValArgGlyValGlyValIleAspPro 350
 DB 2 TATTTCACGGGTGTCAACAGCATGAGATGCGGACATCCGAGGAGGAGGCTTCGATCG 61
 QY 351 AlaTyrMetValIleAspPheGlnLeuMetIleTyrIleGlyAlaAsnSerPheArgThr 370
 DB 62 CCGCTCGTGGTGAAGACTTCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 121
 QY 371 SerHisTyrProTyrAlaGlnGlnValIleMetAspPheAlaAspArgAsnGlyIleVal 390
 DB 122 AGCCACTACCCCTATGCGAGAGAGTATGCGAGTGTGTGACCGCTATGGGATTTGCTC 181
 QY 391 IleAspGluThrProAlaValIleGlyLeuAsnIleAlaLeuMetGlyValSerGluSerGly 410
 DB 182 ATCGATGAGTGTCCCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCT 214
 QY 411 AlaProGlnThrPheThrProAspAlaIleAsnAspIleThrGlnGlnAlaHisLeuGln 430
 DB 215 ---CCGCACTTCTC-----AACAACTTCTCGATCAACACATGCG 256
 QY 431 AlaIleArgGluLeuIleAlaArgAspIleAsnIleAspValIleMetTyrSerIle 450
 DB 257 GTGATGGAAGAGTGTGCTGATGAGCAACCAACCCCGGCTGTATGTGTCTGTG 316
 QY 451 AlaAsnGluProAlaSerHisGlnAspGlyAlaArgGluTyrPheGluProLeuThrAsn 470
 DB 317 GCCAAGACCTCGCTCCCACTAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 376
 QY 471 LeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGlyThrAlaThr 490
 DB 377 CACACCAATCTTGTGACCCCTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 430
 QY 491 TyrGlnLeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArgTyrPheGly 510
 DB 431 TATGACGACAGACAGAGGCTCCGATGTGATGTGATGTGATGTGATGTGATGTGATGTGAT 490
 QY 511 TrpTyrSerGlnThrGlyAspLeuGlnGluAlaGlnAlaIleLeuGlnIleGlnLeuHis 530
 DB 491 TGGTATCAGACATGACGAGGCTGAGTGTGATTCAGCTGCGAGCTGCCACCGATTTGAG 550
 QY 531 GlyTyrGlnGluLeuPheHisArgProIleValIleMetThrGluTyrGlyAlaAspThrLeu 550
 DB 551 AACTGTATAGAGAGTATCAGAGCCATTTATTCAGAGGAGTATGAGACCAAAACGATT 610
 QY 551 AlaGlyLeuHisSerIleLeuGlyLeuProTyrPserGluGluPheGlnValGlnMetLeu 570
 DB 611 GCGAGGTTTCACAGATCCACTCTGATGTTTACATGAGATGATCCAGAAAGTCTGCTA 670
 QY 571 AspMetTyrHisArgValPheAsp-----ArgIleGluSerMetAlaGlyGlnIleVal 588
 DB 671 GACACAGACATCTGGGTCTGATCAAAACCCAGAAATATAGTGTGTGAGAGCTCATTT 730
 QY 589 TrpAsnPheAlaAspPheGlnThrAsnLeuGlyIleIleArgValAspGlyAsnLys 608
 DB 731 TGGAAATTTGCCATTTTCACTGATGACGACGTCACGACGAGAGTGTGCGGGAATTAATA 790
 QY 608 GGIValPheThrArgAspArgLysProLysAlaIleAlaHisSerLeuAlaArgTr 628
 DB 791 GGGGATTTTACCTCGGACAGACAAACAAAGTGAAGGCTTCTTTCGAGAGAGATA 850
 QY 628 pThrSerIleAspLys 633
 DB 851 CTGGAGATTTGCATAA 866

RESULT 14

REFERENCE
1 (bases 1 to 877)
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
AUTHORS
Yi Y., Desai R., Olarte M., Henthorn P. and George A. L.
TITLE
Expressed sequence tags from Canine heart
JOURNAL
Unpublished (2003)
COMMENT
Other ESTs: CH3#042_G03T3
Contact: George AL
Division of Genetic Medicine
Vanderbilt University
529 Light Hall, 2215 Garland Avenue, Nashville, TN 37232-0275, USA
Tel: 615 936 2660
Fax: 615 936 2661
Email: al.george@vanderbilt.edu
Insert Length: 2032 Std Error: 0.00
Seq primer: T7: TAATCGACTCACTATAGCG
High quality sequence start: 39
High quality sequence stop: 859.
Location/Qualifiers

FEATURES
source
1..877
/organism="Canis familiaris"
/mol_type="mRNA"
/db_xref="taxon:9615"
/clone="CH3#042_G03"
/tissue_type="heart"
/cell_type="heart"
/dev_stage="mixed developmental stages (adult, 30 day - 40 day fetal)"
/clone_lib="Canine heart normalized cDNA library in pBluescript"
/note="Organ: heart; Vector: pBluescript; Site 1: 5' of vector NOTI; Site 2: 3' of vector EcorI; Tissue source: dog heart (adult, 30 day - 40 day fetal), right and left atria and ventricle. Dog breed - mixed (beagle, German shepherd, pointer, Irish setter). Library construction: oligo-dT primed"

ORIGIN

Alignment Scores:

Pred. No.: 1.54e-65 Length: 877
Score: 683.00 Matches: 135
Percent Similarity: 62.54% Conservative: 47
Best Local Similarity: 46.39% Mismatches: 87
Query Match: 20.36% Indels: 22
DB: 7 Gaps: 3

US-10-757-093-4 (1-634) x CF406520 (1-877)

```

QY 192 G|YLSARGLI|EGLIN|THYRGLNH|ASPHETYS|ANTYRALAG|YLEU|AARGSER 211
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 46 GGTACTCTTCGTCGACACACTACTTTCCTTCAACTACGCGGCGTCGATCGCCCT 105
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 212 I|E|T|P|L|E|U|Y|S|E|R|V|A|P|R|O|G|L|N|H|I|S|I|E|G|L|A|S|P|L|E|T|H|V|A|L|V|A|L|T|H|S|P|V|A|L 231
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 106 GTGCTCCCTCTACACACACTACTTCACTACGACGACGACGACGACGACGACG 165
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 232 A|S|P|G|L|Y|S|E|R|V|A|L|E|U|L|E|A|S|N|T|Y|R|G|L|V|A|L|G|U|A|L|A|A|S|N|G|L|N|T|H|G|L|N 251
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 166 GACCAAGCACTGGGCTGTGATTCACGATTTTGTCCAGGCGCAGTGAACCTTCCAG 225
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 252 I|L|E|G|L|S|E|R|V|A|L|E|A|S|P|L|Y|A|L|E|V|A|L|A|L|Y|A|L|S|E|R|G|L|Y|A|L|G|L 271
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 226 CTGGAAGTACTCTTCGATGAGGAGGACGAGGCTGCGCCAGGAGGAGGAGGAG 285
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 272 G|Y|H|V|A|L|H|I|E|P|R|O|S|E|R|V|A|L|Y|S|E|U|T|R|P|G|L|N|P|R|O|-----G|Y|A|A 286
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 286 GGGCGGCTGCGAGTGCCCAATGTCACCTCTGTCGCGCCGACTGATGATGACACCCC 345
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 287 A|L|A|T|Y|R|L|E|U|Y|R|G|L|E|U|G|L|V|A|L|S|N|L|E|V|A|L|G|Y|-----S|E|R|S|E|R|G|Y|A|S|P|V|A|L 304
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 346 GCTTACCTGTACTCTCTGAGGTGAGGCTGACTGCGCAGATGCGCGCTGCTGTGCA 405
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 305 A|S|P|H|T|Y|R|A|S|N|L|E|U|A|L|H|T|R|G|Y|V|A|L|A|G|T|H|V|A|L|S|V|A|L|A|G|Y|S|E|R|G|L|N|P|H|E|U 324
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

DB 406 GACTTATACCTCTCCCGTGGGATTCGCACCGTGGCCGCTCACAGAGCCCGACTTCTC 465
QY 325 I|L|E|A|N|G|L|Y|S|P|R|O|P|H|E|T|R|P|H|E|T|H|G|L|P|H|E|G|L|Y|S|H|I|S|G|U|A|S|P|H|T|R|A|L|V|A|L|A|R|G 344
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 466 ATCAACGGGAAACCTTCTATTTCCATGGGGTCAACAACATGAGATGCGCATATCCGA 525
QY 345 G|L|Y|S|E|G|L|Y|H|S|A|S|P|R|O|A|L|Y|T|M|E|T|V|A|L|H|S|A|S|P|H|E|G|L|N|E|U|M|E|T|Y|S|T|R|P|L|E|G|Y 364
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 526 GGGAGGGGCTTTGACTGGCCCGCTGTGTGAAGACTTCAACCTGTGGCGCTGGCGGC 585
QY 365 A|L|A|S|N|S|E|R|P|H|E|A|R|G|H|S|E|R|H|I|S|T|Y|R|P|R|O|T|Y|R|A|G|L|U|G|L|V|A|L|E|U|A|N|L|E|A|L|E|U|M|E|T 384
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 586 GCAATGCTTCCGACACGACCTACCCCTACCGGAGGAGTGAAGCAAGCTTGGCAG 645
QY 385 A|R|G|A|N|G|Y|L|E|V|A|L|I|L|E|A|S|P|L|U|T|H|P|R|O|A|L|V|A|L|G|L|Y|L|E|A|N|L|E|A|L|E|U|M|E|T 404
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 646 CGCTATGGGATGTGTGATGACGAGAGCCCTGTGTGGCATCATGCTGTGTC----- 699
QY 405 G|Y|V|A|S|E|R|G|L|U|S|E|R|G|Y|A|L|P|R|O|G|L|N|T|H|P|R|H|E|T|H|P|R|O|A|P|A|L|I|E|A|S|N|A|P|L|Y|S|E|T|H| 424
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 700 -----CAGACTACAGCAATGTGTCC 720
QY 425 G|I|N|G|U|A|L|A|H|S|Y|S|G|L|A|L|I|L|E|A|R|G|L|U|E|U|L|E|A|A|R|G|A|P|L|Y|S|A|S|H|I|S|A|L|S|E|R 444
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 721 CTGCAACACCACTGAGGTGATGGGAGCTGTGCTGGGATGAAGATACCACTCT 780
QY 445 V|A|L|V|A|L|E|T|T|P|S|E|R|I|L|A|A|S|N|G|L|U|P|R|O|A|S|E|R|H|I|S|G|U|A|S|P|G|Y|A|L|A|R|G|L|U|T|Y 464
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 781 G|T|A|G|C|A|T|G|G|T|C|T|G|T|G|C|C|A|T|A|G|C|C|A|C|T|T|C|T|G|A|A|G|C|C|G|C|T|T|A|C|T|A|C 840
QY 465 P|H|E|G|L|U|P|R|O|L|E|U|H|A|S|N|L|E|U|T|H|A|R|G|L|N|L|E|U 475
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 841 T|T|C|A|A|G|A|G|C|T|A|T|T|G|T|C|A|C|A|G|G|C|C|T|T 873

```

Search completed: March 22, 2005, 21:57:16
Job time : 4490 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model1

Run on: March 18, 2005, 23:42:44 ; Search time 43 Seconds
(without alignments)
1418.638 Million cell updates/sec

Title: US-10-757-093-4

Perfect score: 3354

Sequence: 1 MKFLGLSLSLAAPSIGTP.....RKPKAAASHLRARMTSIDKN 634

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR79: *
2: pir1: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1711	51.0	603	1	GBEGCC
2	1338	39.9	651	2	A26581
3	1335	39.8	648	2	A25047
4	1333	39.4	648	2	A32576
5	1186	35.4	368	2	A85768
6	1186	35.4	370	2	D90919
7	882	26.3	563	2	A72300
8	779	23.2	570	2	C90485
9	524	15.6	237	2	B85768
10	524	15.6	237	2	B85768
11	426.5	12.7	1087	2	F72283
12	408.5	12.2	1307	2	T35944
13	390	11.6	996	2	D86872
14	389.5	11.6	1034	2	T30574
15	382.5	11.4	1014	2	C83990
16	376.5	11.2	1034	2	T30551
17	372.5	11.1	1024	1	GBEC
18	367	10.9	1025	1	TC1266
19	359.5	10.7	755	2	D98482
20	355.5	10.6	1024	2	E90678
21	355.5	10.6	1024	2	A85529
22	353	10.5	1015	2	I39697
23	342	10.2	1026	2	A49750
24	332	9.9	1042	2	E85968
25	332	9.9	1042	2	F91123
26	331	9.9	1060	2	A10201
27	329.5	9.8	897	2	A39405
28	329	9.8	1042	1	GBEC
29	325	9.7	1005	2	T31333

30	323.5	9.6	626	2	A42891	beta-galactosidase
31	311.5	9.3	716	2	JU0275	beta-galactosidase
32	308	9.2	1075	2	T47603	beta-galactosidase
33	287	8.6	1007	2	A30093	beta-galactosidase
34	283.5	8.5	1034	2	A24925	beta-galactosidase
35	261.5	7.8	2228	2	E97942	beta-galactosidase
36	260.5	7.8	2233	2	B95075	beta-galactosidase
37	169	5.0	900	2	T19689	hypothetical prote
38	160.5	4.8	820	2	T37230	probable beta-mann
39	139.5	4.2	448	2	JM0038	beta-glucosidase (
40	139.5	4.2	785	2	H72228	hypothetical prote
41	126.5	3.8	448	2	T48214	endo-1,4-beta-mann
42	126.5	3.8	499	2	A27198	cellulase (EC 3.2.
43	124.5	3.7	499	2	JM0111	cellulase (EC 3.2.
44	124	3.7	891	2	A82755	beta-mannosidase p
45	123.5	3.7	403	2	A84592	(1-4)-beta-mannan

ALIGNMENTS

RESULT 1
GBEGCC
beta-glucuronidase (EC 3.2.1.31) uida - Escherichia coli (strain K-12)
N:Alternate names: beta-D-glucuronoside glucuronosylhydrolase; gusa protein
C:Species: Escherichia coli
C>Date: 30-Jun-1988 #sequence revision 05-Dec-1997 #text_change 09-Jul-2004
C/Accession: C64918; 153717; A26487; S4355
R/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Coi
.A.; Rose, D.J.; Mau, B.; Shaoh, Y.
A>Title: The complete genome sequence of Escherichia coli K-12.
Science 277, 1453-1462, 1997
A/Reference number: A64720; MUID:97426617; PMID:9278503
A/Accession: C64918
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-603 <RES>
A/Cross-references: UNIPROT: P05804; GB: A8000257; GB: U00096; NID: g178798; PIDN: AAC74689.1
A/Experimental source: strain K-12, substrain MG1655
R/Schlaman, H.R.; Risseuw, E.; Franke-van Dijk, M.E.; Hooykaas, P.J.
Gene 138, 259-260, 1994
A>Title: Nucleotide sequence corrections of the uida open reading frame encoding beta-gl
A/Reference number: 153717; MUID:94171050; PMID:8125312
A/Accession: 153717
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-603 <RES>
A/Cross-references: GB: S69414; NID: G545893; PIDN: AAB30197.1; PID: G545894
R/Jefferson, R.A.; Burgess, S.M.; Hirsh, D.
Proc. Natl. Acad. Sci. U.S.A. 83, 8447-8451, 1986
A>Title: Beta-glucuronidase from Escherichia coli as a gene-fusion marker.
A/Reference number: A26487; MUID: 87041472; PMID: 3534890
A/Accession: A26487
A/Molecule type: DNA
A/Residues: 1-419; 'VHGNIS', 427-603 <JEF>
A/Cross-references: GB: M14641; NID: G868017; PIDN: AAA68923.1; PID: G868020
R/Punt, P.
submitted to the EMBL Data Library, April 1994
A/Reference number: S43555
A/Accession: S43555
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1, 'V', 3-603 <PUN>
A/Cross-references: EMBL: Z32701; NID: g475168; PID: g475169
C/Comment: This acid hydrolase catalyzes the cleavage of a wide variety of beta-glucuron
C/Genes:
A:Gene: uida
A/Map position: 36 min
C/Function:
A/Description: catalyzes hydrolysis of beta-D-glucuronoside to D-glucuronate and alcohol.
A/Pathway: starch and glucose metabolism
C/Superfamily: beta-glucuronidase
C/Keywords: glycosidase; hydrolase

Query Match 51.0%; Score 1711; DB 1; Length 603;
Best Local Similarity 55.0%; Pred. No. 6,1e-111;
Matches 330; Conservative 93; Mismatches 161; Indels 16; Gaps 9;

QY VRPQRSSRELVLDGLMKFAL---ASGNDTQNPVTAFLPKGLBCEPVPASVNDIFISSE 96
DB LRRVEFTPEIKKLDGLMAFSLDRENGCIDQ--RWMESALQSRALAVAGSFNDQPADAD 59
QY 97 IHDFGVWVYVYQREVIVPKGMSQERYLVRAESATTHGRIVYNNRLVAEHGQGYTPEADYT 156
DB 60 IRNVAGNVWYQREVEFIPKGMAGORIVLRFDATVHYGVKVVNNQVEVHEQGYTPEADYT 119
QY 157 ELVAPBEKRLITGVNNELTHETIPPKTITTGNAATGRVQYQHPFNAYAGLARSITWYS 216
DB 120 PVLAKGSVRIVVCNNELNMQITPPGAWIT--DENGKKQASVHFHFNAYAGHRSVMYLT 178
QY 217 VPOQHIDITVYVDV--DGDNGLINEVEVANAQTTQIOISVIDEIGAIVAASAGQGTV 274
DB 179 TENVWDDITLVTHVADCNHASVDNQV--VAN---GDVSELRDADQOVVATGQGTSTL 234
QY 275 TTPSVYKLMOPGAAYLYQLQVNIYVSSGVDVDTYNLATGVRTVKVAGSOFILNGKEPFYTG 334
DB 235 QVNVPELMPQGGGYELTLCV--TAKSQTECDITPIPLRVGIRSAVAVGEGRLNHRKFFYTG 292
QY 335 FKGHELTAVRGSGHPVAVWHDFOLMKMTGASFTSTSHPYAEVMDPADRGIVYIDET 394
DB 293 FGRHEDADRGRGFDVLWHDHALMDWIGANSYRSHSPVAEMLDWADEGIVYIDET 352
QY 395 PAVGNTIAL-MGVSESGAP-QTFEPDAINDKQEAHKAIRFLIARDKNHASVVMASIAN 452
DB 353 AAVGNLSLGIFPAGNPKKELYSEAVNGEIQOHLQAIKELIARDKNHSVVMASIAN 412
QY 453 EYASHEDGARFEFPELITNLTRQLDPTPTTANVTATYQLDRIISDLFVDSICNRYFGWY 512
DB 413 EBDTPQGRAREVFAPLAETATKLDPTPTTCVNAWFCDAHDTTISDLFVDCINRYGYWY 472
QY 513 SOTGLEBEAALTEKELHGMQEKFRPIVMEYEGADTLAGHSITLGLPMSEEPQVQULMD 572
DB 473 VQSGLETRKEKLEKELAMQEKHOPITITRYGADTLAGHSMTDWMSEBYQCAMLDM 532
QY 573 YHVFDRIESNAGEHVWNPADFQTNIGIIRVDGNKKGVFTDRKPKAAHSLARAWTSID 632
DB 533 YHVFDRVSAVYGEQVWNPADFATPSQGLIRVGNKKKGLFTDRKPKSAAPFLQKRWTCMN 592

RESULT 2
A26581
beta-glucuronidase (EC 3.2.1.31) precursor - human
CSpecies: Homo sapiens (man)
CDate: 05-Oct-1988 #sequence revision 05-Oct-1988 #text_change 09-Jul-2004
CAccession: A26581; A40337; A24983; A26538
RShimizu, A.; Kyle, J.W.; Miller, R.D.; Hoffmann, J.W.; Powell, P.P.; Grubb, J.H.; Sly,
Proc. Natl. Acad. Sci. U.S.A. 84, 685-689, 1987
A>Title: Cloning, sequencing, and expression of cDNA for human beta-glucuronidase.
A:Reference number: A26581; MUID:87118233; PMID:3468507
A:Accession: A26581
A:Molecule type: mRNA
A:Residues: 1-651 <OSH>
A:Cross-references: UNIPROT:P08236; GB:IM5182; NID:g183223; PIDN:AAA5261.1; PID:g183233
A:Experimental source: placenta
R:Shupley, J.M.; Miller, R.D.; Wu, B.M.; Grubb, J.H.; Christensen, S.G.; Kyle, J.W.; Sly,
Genomics 10, 1009-1018, 1991
A>Title: Analysis of the 5' flanking region of the human beta-glucuronidase gene.
A:Reference number: A40337; MUID:92009900; PMID:1916806
A:Accession: A40337
A:Molecule type: DNA
A:Residues: 1-70 <SH1>
A:Cross-references: GB:M65002; NID:g183706; PIDN:AAA5262.1; PID:g183707
R:Guise, K.S.; Korneluk, R.G.; Wayer, J.; Lambornwah, A.M.; Quan, F.; Palmer, R.; Ganeschow
Gene 34, 105-110, 1985
A:Reference number: A24983; MUID:85232043; PMID:3924735
A:Accession: A24983

A:Molecule type: mRNA
A:Residues: 520-585 <GUI>
A:Cross-references: GB:MI0610; NID:G183704; PIDN:AAA52621.1; PID:G183705
R:Tomatsu, S.; Fukuda, S.; Sukegawa, K.; Ikeda, Y.; Yamada, S.; Yamada, Y.; Sasaki, T.; (Am. J. Hum. Genet. 48, 89-96, 1991
A:Title: Mucopolysaccharidosis type VII: characterization of mutations and molecular heterogeneity
A:Reference number: A36538; MUID:91090114; PMID:1702266
A:Accession: A36538
A:Molecule type: mRNA
A:Residues: 378-385, 616-621, 643-651 <TOM>
C:Genetics:
A:Gene: GDB:GUSB
A:Cross-references: GDB:120025; OMIM:253220
A:Map position: 7q22-7q22
C:Superfamily: beta-glucuronidase
C:Keywords: glycoprotein; glycosidase; homotetramer; hydrolase; lysozyme
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-651/Product: beta-glucuronidase, placental #status predicted <MAT>

Query Match: 39.9%; Score 1338; DB 2; Length 651;
Best Local Similarity 44.3%; Pred. No. 5,7e-85;
Matches 274; Conservative 98; Mismatches 199; Indels 48; Gaps 12;

Qy	42	PORTSSSEIIVNIGISLMEKALASGLINDT-----ACQWTPYLKRG---LECPRPASVYNIF	92
Dd	30	POESPSTBECKELDLSNFS--RADPSNRRRGFEEOQYRRRLPESGPTQMDPVPSFNDIS	87
Qy	93	ISREIHHDVGVVYVQREIVIVPKGMSOE---RYLVARESAATHGRITVYNNRLAEHVAGYT	144
Dd	88	QDMRLRFHVGVMVYEREVLIPERMTQDLTRVYLIRIGSHSAIYVWAGVDTLHNEGYL	14
Qy	150	PFEDVTVELVAFG---EKFRLLTGANNELTHETIIPRGT-----TTGNATGKGIOTYQHD	20
Dd	148	PFEDIDSNLVQVPERBRLRTITAINNLTLPRTLPRGTIQLYLLDTSKPKGYFQNTYED	20
Qy	202	FYNAGIARSIIMELYSVPOQHIODITVTVTDVGDNGILINVEVANAQTGGQIDISVDEBDG	26
Dd	208	FFNVAGIGRSLVLTPTTPTTYIDITVTTTSVEQDSGLVYQISVKSNNFLKLEVRLLDAEN	26
Qy	262	AIYAKASGAQGTVTIDSVKLMOP-----GAAYLYOLQVNIYVG--SSGDVDTYMLATGVR	314
Dd	268	KVYANGTGTGGOLKVPQSVLIMPEYLMHERRPALYLSLEVQLTQTSIGPVSDFYTLVPGIR	32
Qy	315	TVYKASGOEFLINKRPYFTFGFNGHRTAARGSGNDPAIYVNDPOLMKRIGANSRPTSTYR	37
Dd	328	TVATTSQFLINKRPYFTFGVNNHEDADIRGKGFPMPLLVQDFNLRLVGNAAFRSTSYR	38
Qy	375	YAEVMDFADRNQIIVIDETPAVGLIATLMGSESGAQPTQETPRADINQTOEAKQAIRE	434
Dd	388	YAEVVMQCMCRYSIIVIDECPGIGAL-----IQFF-----NNVSLHHNQWEE	43
Qy	435	LIADKKNHSAVVMWSIANEPRASHEDGAREYEPRLTNLTROLDPTRPITFANVGTATYQLD	494
Dd	433	VVRDKDKHRAVVMWSVANEPRASHLESAGYLLTMVAIHTKSLDPSRPVTF--VSNSNYAAD	49
Qy	495	RISDLFPVSCINNYFCWYSQTDGLEBAALKEKELHMGQEKPHRPIVWTEYCATLACGLH	55
Dd	491	KGAPYVDVICMSYKWHYNDYGHLELIQDLATQRENNYKKYQKRTIQSEYGAEITLIGFH	55
Qy	555	SLIGLPMSEEPQVOMIDMYHVRFD--RIESNAGEFVNNVAFDPOTNLGIIRVGNKKKCVFT	612
Dd	551	QDEPLFMTEBYEQKSLLEQYHLGLDQKRRKYVVGELINMFADPMTEQSPTRLVGNKKCIFT	610
Qy	613	RDRKPKAAHSLPARKTSI 631	
Dd	611	RORQPKSAAFLLRERYWKI 629	

QY 42 PORTSHELVDGLMKFAL--ASGLINDTAQPW-TAPLRK---LECPVAPASYNIDIFIS 94
 Db 30 PKSPRSRLKALDGLMHFRADLSNNRLOQGFEOQYRQPLNESGVLMDPSSPFNDITOE 89
 QY 95 REIHDHGWYVYQREVIVPKGMSOE--RLVRAESKTHGGRIVANNRLVAEHVGGTTF 151
 Db 90 AALRDFGWMYEREAALPRRWTODTDMRVLRNSAHYAVVWVNGIHVHEEGGLPFP 149
 QY 152 EADVELIAPG--EKFPLTIGVNNELTHEFTIPPKI-----TTGNATGKRIQTYOHDFYN 204
 Db 150 EADISLVOSGPELTKCTITAINNTLTPHLPFGTIVYKTDTSMPRGIVQDTSDFEN 209
 QY 205 YAGLANSIMLVSPQOHIODITVTVDVGDNGELINTEVEVANOQTGOIQISVIDEGAIY 264
 Db 210 YAGLHSSVVLVYTPPTVYIDITIVITVNEODIGLVYTWISVQSEHGFLEVLDEGGKV 269
 QY 265 AKASGAGTITISVLMOP-----GAAYLYOQVNI VSSGSDVDTYNLATGVRVYKA 319
 Db 270 AHGTGNOGQLOVPSANLMMWYLMHEHPAYWSLEVKT--TTESVDTYTLPIGIRVAAYT 328
 QY 320 GSQFLINGKPEYFGEGKHEDTAVRGKHPAYMVHDFOLMKWIGANSFRTSHYPAEEV 379
 Db 329 KSFFLINGKPEYFGVGNKHEDSDIRGKFPWPLVVDNLRLKANSFRTSHYPSSEV 388
 QY 380 MDPADNRGIVYIDETPAVGNLALMGVSSGADQTFPPDAINDKTOBAHKAQIRELIAD 439
 Db 389 LQCDHYGVYVIDECBGVIVL-----POSGNESIRH-----HLEVEEELVRSD 433
 QY 440 KHNASVYVMSIANEPASHEDGAREFEPLTNLTROLDPRPIFANVGTATYOLDISDL 499
 Db 434 KNPBAVYVMSVANSPPSALKPRAVYFETLTHTKALDLTRPVF--VSNKADYADICAPY 491
 QY 500 FVDSICINRYFGWYSGTQDLEAEALAEKELHGWQEFHRIPTVTEGADTLAGHSILGL 559
 Db 492 VDVTCNSYYSWYHDVGHLEVIQPOLNSGFENYKTHQKRIQSEVGAADLPIHDDPR 551
 QY 560 PMSEEFQVOMLDMYHRYVD--RIESMAGHEVWNPADPOTNLGIIRVDGNKKGVFTDRKP 617
 Db 552 MFSEEFQVOKLVENYHSLDQKRKEVYVGLIWNFADFMNTOSPLRAVIGNKKGIFTROBP 611
 QY 618 KAAHSLRLARWTSI 631
 Db 612 KTSAPFLRERYKRI 625

RESULT 5
 A85768
 partial beta-D-glucuronidase [imported] - Escherichia coli (strain O157:H7, substrain ED
 C/Spectes: Escherichia coli
 C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
 C/Accession: A85768
 R./Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.U.; Mayhew
 Iller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; DiMantano, E.; Potamouotis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A./Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A./Reference number: A85480; WUID:21074935; PMID:11206551
 A./Accession: A85768
 A./Status: preliminary
 A./Molecule type: DNA
 A./Residues: 1-368 <STO>
 A./Cross-references: UNIPROT:Q8X671; GB:AE005174; NID:g12515602; PIDN:AGS6605.1; GSPDB:G
 A./Experimental source: strain O157:H7, substrain EDL933
 C./Genetics:
 A./Gene: uidA_1

Query Match 35.4%; Score 1186; DB 2; Length 368;
 Best Local Similarity 63.3%; Pred. No. 8.6e-75;
 Matches 224; Conservative 41; Mismatches 85; Indels 4; Gaps 3;

QY 281 LMOPGAAYLYOQVNI VSSGSDVDTYNLATGVRVYKAVGSOFLINGKPEYFGKHED 340
 Db 6 LMOPGEGYLYLV--TAKSRTECDIYPLRVGIRSAVAGGQFLINHKPFYFTGGRHED 63

QY 341 TAVRGKHPAYMVHDFOLMKWIGANSFRTSHYPAEEVMDPADNGIYVIDETPAVGIN 400
 Db 64 ADLRGKGFNVLMVHDMALMDWIGANSYRTSHYPAEEMLDWADDEGIVYIDETAVGN 123
 QY 401 IAL-MGVSSGAP-QTFPPDAINDKTOBAHKAQIRELIARDQNHASVWMSIANEPASHE 458
 Db 124 LSLGIFPEAGNPKELYSEEAANGETOAHLOAIKELIARDKNHSVWMSIANEPDTRP 183
 QY 459 DGAREYFELTNLTROLDPRPIFANVGTATYQDRISDLFDVSCINRYFGWYSGTDL 518
 Db 184 OGAREYFAPLABATKLDPTTRPITCVNVWPCDAHTDTLSDFVCLNRYTGMVQSGDL 243
 QY 519 EBAEALAEKELHGWQEFHRIPTVTEGADTLAGHSILGLPMSEEFQVOMLDMYHRYVD 578
 Db 244 ETAEKYLEKELLAMOEKHLQPIITIEYVDTLAGHSMTVDWMSSEYQCAWLDMYHRYVD 303
 QY 579 RIESMAGHEVWNPADPOTNLGIIRVDGNKKGVFTDRKKAALSLRARWTSID 632
 Db 304 RVSAVAGQVWNPADPATSQGITRVGNKKGIFTDRKPKSAAPFLQKRWGMN 357

RESULT 6
 D90919
 beta-D-glucuronidase [imported] - Escherichia coli (strain O157:H7, substrain RIMD 050991
 C/Spectes: Escherichia coli
 C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
 C/Accession: D90919
 R./Hayashi, T.; Makino, K.; Onishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
 gagaara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
 DNA Res. 8, 11-22, 2001
 A./Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom
 A./Reference number: A99629; WUID:21156231; PMID:11258796
 A./Accession: D90919
 A./Status: preliminary
 A./Molecule type: DNA
 A./Residues: 1-370 <HAY>
 A./Cross-references: UNIPROT:Q8X671; GB:BA000007; PIDN:BA835747.1; PID:g13361791; GSPDB:G
 A./Experimental source: strain O157:H7, substrain RIMD 0509952
 C./Genetics:
 A./Gene: EC62324

Query Match 35.4%; Score 1186; DB 2; Length 370;
 Best Local Similarity 63.3%; Pred. No. 8.7e-75;
 Matches 224; Conservative 41; Mismatches 85; Indels 4; Gaps 3;

QY 281 LMOPGAAYLYOQVNI VSSGSDVDTYNLATGVRVYKAVGSOFLINGKPEYFGKHED 340
 Db 8 LMOPGEGYLYLV--TAKSRTECDIYPLRVGIRSAVAGGQFLINHKPFYFTGGRHED 65
 QY 341 TAVRGKHPAYMVHDFOLMKWIGANSFRTSHYPAEEVMDPADNGIYVIDETPAVGIN 400
 Db 66 ADLRGKGFNVLMVHDMALMDWIGANSYRTSHYPAEEMLDWADDEGIVYIDETAVGN 125
 QY 401 IAL-MGVSSGAP-QTFPPDAINDKTOBAHKAQIRELIARDQNHASVWMSIANEPASHE 458
 Db 126 LSLGIFPEAGNPKELYSEEAANGETOAHLOAIKELIARDKNHSVWMSIANEPDTRP 185
 QY 459 DGAREYFELTNLTROLDPRPIFANVGTATYQDRISDLFDVSCINRYFGWYSGTDL 518
 Db 186 OGAREYFAPLABATKLDPTTRPITCVNVWPCDAHTDTLSDFVCLNRYTGMVQSGDL 245
 QY 519 EBAEALAEKELHGWQEFHRIPTVTEGADTLAGHSILGLPMSEEFQVOMLDMYHRYVD 578
 Db 246 ETAEKYLEKELLAMOEKHLQPIITIEYVDTLAGHSMTVDWMSSEYQCAWLDMYHRYVD 305
 QY 579 RIESMAGHEVWNPADPOTNLGIIRVDGNKKGVFTDRKPKAALSLRARWTSID 632
 Db 306 RVSAVAGQVWNPADPATSQGITRVGNKKGIFTDRKPKSAAPFLQKRWGMN 359

RESULT 7
 A72300
 beta-D-glucuronidase - Thermotoga maritima (strain MS8)

A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-237 <HAY>
A:Cross-references: UNIPROT:Q9AHJ5; GB:BA000007; PIDN:BAH35748.1; PID:g13361792; GSPDB:Q
A:Experimental source: strain 0157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: EC82325

Query Match 15.6%; Score 524; DB 2; Length 237;
Best Local Similarity 44.9%; Pred. No. 4.4e-29;
Matches 105; Conservative 47; Mismatches 74; Indels 8; Gaps 5;

QY 40 VRPQRTSSRELVDLGMKFPAL-ASGLNDTAQPTAPLPKGLCECPVASYNDIFISREIH 98
DB 2 LRPEVETPREIKKL DGLMAFSLDRENGCIEQRWESALCESRAIAVPSGSDQFADADIR 61
QY 99 DHVGMVYVYQREVIYVPGKMSQERYIVRAESATHHGRIVYNNRLVAEHVGGYTPFEADVTEL 158
DB 62 NYGVANWYQREVIYVPGKMGQRIVLRFDAVTHGKVVANNQEVMEHGGYTPFEADVTEL 121
QY 159 VADGEKFRLLTGNNELTHEITPPKITTGNATGRIQTYQHDFFVYAGLARSIMLYSVP 218
DB 122 VIAGKSVRITVCNNELMQITIPPQWVIT-DENGKKQSYFHDFFVYAGIHRSVMLYTPP 180
QY 219 QOHIDITVTVTDV--DGDNGLIYEVYVANNQTTGOIQISVIDEDGAIYAKASGA 270
DB 181 NTWVDITVTVTHVAQDCNHASVDMQV-VAN--GDVSVELRDADQVVVATGGA 230

RESULT 10
B85768
partial beta-D-glucuronidase [imported] - Escherichia coli (strain 0157:H7, substrain ED
C/Species: Escherichia coli
C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C/Accession: B85768
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-237 <STO>
A:Cross-references: UNIPROT:Q9AHJ5; GB:AE005174; NID:g12515603; PIDN:AA656606.1; GSPDB:Q
A:Experimental source: strain 0157:H7, substrain EDL933
C:Genetics:
A:Gene: uidA_2

Query Match 15.6%; Score 524; DB 2; Length 237;
Best Local Similarity 44.9%; Pred. No. 4.4e-29;
Matches 105; Conservative 47; Mismatches 74; Indels 8; Gaps 5;

QY 40 VRPQRTSSRELVDLGMKFPAL-ASGLNDTAQPTAPLPKGLCECPVASYNDIFISREIH 98
DB 2 LRPEVETPREIKKL DGLMAFSLDRENGCIEQRWESALCESRAIAVPSGSDQFADADIR 61
QY 99 DHVGMVYVYQREVIYVPGKMSQERYIVRAESATHHGRIVYNNRLVAEHVGGYTPFEADVTEL 158
DB 62 NYGVANWYQREVIYVPGKMGQRIVLRFDAVTHGKVVANNQEVMEHGGYTPFEADVTEL 121
QY 159 VADGEKFRLLTGNNELTHEITPPKITTGNATGRIQTYQHDFFVYAGLARSIMLYSVP 218
DB 122 VIAGKSVRITVCNNELMQITIPPQWVIT-DENGKKQSYFHDFFVYAGIHRSVMLYTPP 180
QY 219 QOHIDITVTVTDV--DGDNGLIYEVYVANNQTTGOIQISVIDEDGAIYAKASGA 270
DB 181 NTWVDITVTVTHVAQDCNHASVDMQV-VAN--GDVSVELRDADQVVVATGGA 230

RESULT 11
F72283
beta-galactosidase - Thermotoga maritima (strain MSB8)

C/Species: Thermotoga maritima
C/Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C/Accession: F72283
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
R.Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hickey,
Garrett, M.W.; Stewart, A.W.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.

Nature 399, 323-329, 1999
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1087 <ADN>
A:Cross-references: GB:AE001776; GB:AE000512; NID:g4981740; PIDN:AAU36268.1; PID:g4981744
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TML193
C:Superfamily: beta-galactosidase

Query Match 12.7%; Score 426.5; DB 2; Length 1087;
Best Local Similarity 24.2%; Pred. No. 2.7e-21;
Matches 157; Conservative 107; Mismatches 231; Indels 153; Gaps 28;

QY 49 ELVNDLGLMKFPALASGLNDTAQPTAPLPKGLCECPV--KGLCECPV-- 84
DB 41 ERLSINGWNRFLFAK-----NPEVPEDPFSEKFDSDNMBEIEVPSNWKGGYKPIYT 94
QY 85 ----PASYNDIFISEIHQVYQREVIYVPGKMSQERYIVRAESATHHGRIVYNNRL 140
DB 95 NVVYVPEPVPFPVKKD--DNPTGV-YRKMIEIPEMFKEFLHPEGVRSFFYLMVNNCK 151
QY 141 VAHVGGYTPFEADVTELVAPEKFRLLTGNNELTHEITPPKITTGNATGRIQTYQH 200
DB 152 IEFKSDSCPPAFRLTDVLRPK-----NLITVEVL--KWSDS-----YLEDD 193
QY 201 DRYNATGLARSLWYVYQOHIDITVTVTDVGD--NGLINVEYVAN--QTGQIDIS 255
DB 194 DMMWVAGIYRDVYVLAALPFRHIDVFRVTDLDENYRNKIFLDVWRNLGSEEEKDLVY 253
QY 256 VIDDEG---AIYAKASGAQGV-----TIPSVKLMQPGAAVLYQVNIIVGSSGDVVDY 307
DB 254 LITPDDEKTLVKEIVKPEDRVLSFAFDVKPKKSAETPHLYVAKLKL--GB--DEK 307
QY 308 NLATGVYVYVYQSOFLNGKPEYFTGPKKHEDTAVRGKHDPAYVYHDFOLMKVIGANS 367
DB 308 KYNFGRKRIEIKDGLTFNGKELYIKGVNRHEDPDGRNAVVERMIDIDIKLMKOHNINT 367
QY 368 FRTSHYVYAEVMDPADRNGIYVIDETAVGINILMGVSESGAPOTFPD--AINDKT 424
DB 368 VRTSHYVYQTKWYDLCDFYGLYVIDEA---NIESHGID-----MDPEVTLANKREW 415
QY 425 OEAHKQALREILARDKNHASVVMWSIANEPASHEGAREYEPRLTNLTROLDPTREPTFA 484
DB 416 EKAFHPRIRKAMBRDKNHSSIIIFMSIGNAG--DSVN--FEKALMTIKRDNLTLYHE 470
QY 485 NVGTA---TYOLDRIISLFDVSCINRYFGWYSQTGDEEAPALAEKELHGOEKFNPI 540
DB 471 --GTRRGSYVYDVVSLYPRKMDIL-----LEYAKSKREK-----PF 506
QY 541 VMTGYADLTLAGHSLGLPMSEEPYQWLDMYHYFDCIESMAGEHYVNFAD----- 593
DB 507 IMCEYA-----HAAGN-----SVGNLKDMDVIEIKYPLHGGCTIDMDWDGIRKDD 552
QY 594 -----FOTNLGIRVDGN--KKGVFTREDRKPKAAHSLRARTSI 631
DB 553 ENGRERWYAGDGRDPNDGNCINGVLPDRTPPELLEVKKVQNV 600

RESULT 12
T35944
probable beta-galactosidase - Streptomyces coelicolor
C/Species: Streptomyces coelicolor
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004

C/Accession: T35944
R/Seeger, K.V.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, January 1999
A/Reference number: Z21551
A/Accession: T35944
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1307 <SEB>
A/Cross-references: UNIPROT:Q9ZBF6; EMBL:AL035206; PDB:CAA22766.1; GSPDB:GN00070; SCOP
A/Experimental source: strain A3(2)
C/Genetics:
A/Genes: SC0EB:SC9B5.24

Query Match 12.2%; Score 408.5; DB 2; Length 1307;
Best local similarity 22.9%; Pred. No. 6.5e-20;
Matches 149; Conservative 80; Mismatches 250; Indels 171; Gaps 21;

17 LGTPAARHPFRNEMTOHEOPLIKVRPORTSSRELVLMDGLMKFALASGLNDTAQ-----70
71 LGTP-----PHATLTPLADVEQALAGDRANSPYRLSLDGTWKFAVADRPEDROADPHRTD 126
71 -----PW-TAPLR-----GLCEPV-----PASYNDIFISREI 97
127 VDDGDGTLTPVPSWOLHGHDPILYNTIYPWPGNGGSEPPQPAAPFTYVNP-----180
98 HDHGVWVYQREVIYVPGMSQERYLVASATHHGRIYVNNRLVAEHVGYTFPEADYTE 157
181 -----GQIRRTITVREDMGRKTPFLFBGVKSANHYWINGELAGDEDSYTTSEIDVTD 234
158 LVAPGEKFLTLTGNNELTHETIPKQITTGATGKRIQTYQH-----DFYVYAGL 208
235 LKPG-----TNOIA-----VEYRYSDGDMDMDQIRLSGI 267
209 ASIMWYSPQOHIQITVTVD-----GNGLINEVEV-----A 245
268 FRSVHLVSTPAVNLRFKLTDPDDTYTGALVTSVRAVYGGGGRYVETOLYDARG 327
246 NOTGQIQISVIDEDAIYVAKSAGQVTPISVKLMQPAAYLYOLQVNIYVSSGVDV 305
328 HAVMSRPLQOPVUGGABAGEDVTVAARAVKPRKLSAHPRLYTLVLRPAPAGVTE 387
306 TYNLATGVTVKVASQFLINGKRFYFTFGKHEDTAVRGKGDPAVYVDFOLMKWIGA 365
388 TLSHRVGFREFALKDGLMRNGEPVSIKGNRHEHMPARGTALTREDMVEDIKIKRMN 447
366 NSRRTIHYPAEVMPPADNGIVVIDETPAVGLNTALMGVSSGAPQTFTPAINDKTQ 425
448 NSVRTSHYPNNPYMLADEYGLYVDET-----NLETHGIRGE-----YFPGDHPDWT 496
426 EAHKQIRELIARLAKHNVMSIANEPASHEDGAREVEPLNLTROLDPRTTFAN 485
497 ACYARQ-QNVNRDKHNSAVIMSLGNEA-----GGGSTSNAHWIRSDTDRLVQYEG 550
486 VGTATYQLDR--ISDLDVSCINRYFGWYSQTDLEBAEALKEKELHGMQEKFRPIVMT 543
551 D-----DRPGVSDI-----RSEWYDSQVRVGRKADISDT--RYVVM 586
544 EYGAOTLAGLHSLGLPWESEFOVQMLDMYHVFDRIESNAGHEVNNPAD 593
587 EYA-----HSMGN-----STGNFKKTYWDVRRYVLOGGIMDFVD 622

RESULT 13

beta-galactosidase (EC 3.2.1.23) [imported] - Lactococcus lactis subsp. lactis (strain
C/Species: Lactococcus lactis subsp. lactis
C/Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C/Accession: D86872
R/Bolotin, A.; Wincker, P.; Manger, S.; Jallion, O.; Malarme, K.; Weisenbach, J.; Ehrlich
Genome Res. 11, 731-751, 2001
A/Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s
A/Reference number: A86625; MUID:2125186; PMID:11337471
A/Accession: D86872

A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-996 <STO>
A/Cross-references: GB:AE005176; PID:g12725024; PDB:AAK06078.1; GSPDB:GN00146
A/Experimental source: strain IL1403
C/Genetics:
A/Genes: lacZ
C/Superfamily: beta-galactosidase
C/Keywords: glycosidase; hydrolase

Query Match 11.6%; Score 390; DB 2; Length 996;
Best local similarity 22.9%; Pred. No. 8.2e-19;
Matches 138; Conservative 96; Mismatches 238; Indels 130; Gaps 23;

49 ELVNDLGMKFKALASGLNDTAQPTAPLRKGLCECPVPSYNDIFISRE--HHDHGVWY 105
34 EKSLINGLNFHDFRSISDVPKWMLTSTKTEIIVPSWQIEFKKQSDVPIYNTVPI 93
106 YOREVIVPKG--WSQERYLVASATHHGRIYVNNRLVAE-----HVGST-----149
94 PIQPPYVPEANPVGARSRYFDITKEWLESQHVLTPBGOSAFHFLNGYGGYSDSRL 153
150 PFEADVTELVPGEK-----FLLTGNNELTHETIPKQITTGATGKRIQTYQHDFY 203
154 PAFEDISNLAKEGQCLKYLVRFS-----KVL-----YFEDQDMW 189
204 NVAGLARSTWYSPQOHIQITVTVDGNDGLIYVE--VANOTGQIQISVIDEDG 261
190 RMSGIFRSVNLQWLPNNYLLDFSIKTDLEDLDFAVNKLQAVAKNIDDALEKLYD-DE 248
262 AIVAKSAGQVTPISVKLMQGAAYLYOLQVNIYVSSGVDVDTYNLATGVTVKVAS 321
249 QLIGCHGPDALGVNPNKLSDEIPLYRLLETLMDRGAHVHKKTKIKIKIAIEKG 308
309 QLKINKKALLVGVNGHEFTPEHGYVSEVMIKDKLMKEHNFNVRCSHPYNDSRWYE 368
382 FDRNGIVVIDETPAVGLNTALMGVSSGAPQTFP--DAINDKTQ-EAHKQIRELIAR 438
369 LCEYGLYVWDE--NLETHG--TWNRLTNDPTYLPLMSEVTRVMR 414
439 DKHNSVMSIANEPASHEDGAREVEPLNLTROLDPRTTF-----ANYGATYQL 493
415 DRNHPSTIIISLGN-----ESGYSNHQALYDMCKSFDSRPHYVGGDDASNG-ATDAT 468
494 DRISDLF--DVSICINRYFG--WYSQTDLEBAEALKEKELHGMQEKFRPIVMTYGA 547
469 DICPMYARVDSPIVAPYSLKTMGVGE-----NRPILCEYAN 509
548 D--TLAGLHSLGLPWESEFOVQMLDMYHVFDRIESNAGHEVNNPADPQTNLGIIRD 604
510 DMGNSLGG-----FGKYWAQFREIDRLQGGIFMDVD--QGLLK-D 547
605 GN 606
548 GN 549

RESULT 14

beta-galactosidase - Bacillus megaterium
C/Species: Bacillus megaterium
C/Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C/Accession: T30574
R/Strey, J.
submitted to the EMBL Data Library, September 1997
A/Reference number: Z20870
A/Accession: T30574
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1034 <STR>
A/Cross-references: UNIPROT:O52847; EMBL:AJ000733; PDB:CAA04267.1

C:Genetics:
A:Gene: bgam
C:Superfamily: beta-galactosidase

Query Match	11.6%;	Score 389.5;	DB 2;	Length 1034;
Best Local Similarity	21.0%;	Pred. No. 9.4e-19;		
Matches 154;	Conservative 104;	Mismatches 261;	Indels 213;	Gaps 28;
QY	14	APSLGTFAARHPPRNMTQHEOPLIKVRPQRTSRELVN-----LDGLMKFALAS	63	
DB	12	APANGTPEKNNNPETQLNRSKHALMRYQYEEALKNDKRSVYVYQSLNGSMYFHPAE	71	
QY	64	GLNDTAQPTAP-----LP-----KGLGEC-----VPASY	88	
DB	72	NADGRVKNFFAPFSEYKMDISVPSHMOLOGDYDPOYTNVYVWVNELEPPAPYKY	131	
QY	89	NDIFISREIHDHGWYVYQREYVVPKWSGERLVAAESATHHGRILYNNRLVAEHVGY	148	
DB	132	NPV-----GQYVFTTPKSEMDQPVYISFGVESAFYVWINGEYVGSSEDF	179	
QY	149	TPEEADVTLELVADGEKRLTIGVNNELTHETIPPGKITTGATGKRIQTYQ-----	199	
DB	180	TPAEFDITSLVLOGEN-----TIAV-----EYVRMSDASMLBD	212	
QY	200	HDFVYVAGLARSIMLYSVPOOHIDITVTVDVG--DNGILNVEEVANQTTGOIQ-----	253	
DB	213	QDFWRMSGIFRDVYLYSTPQVHIYDFSVRSLSNNYEDGELVSADILNFEHDTDLTF	272	
QY	254	-ISVIDEDGALVAKAS-----GAQGYT-----IPSVKLMGCAAVLYQLOVNIIVGSSG	301	
DB	273	EVMILYDNAQOVLQAPLOTLSVSDQRTVSLRTHIKSPAKMSAESPLNLYTLVLSLKAAG	332	
QY	302	DVVDVYVNLATGRTVTVKVASGQFLNGKPYFTFGKHEDTAVRGK-GHDPAYVHDPQLM	360	
DB	333	SIIFTECKGQGFREFEIKNGIMTNGKRIYLRGVNRHEFDSYKGRAGITREDMIHILLM	392	
QY	361	KWIGANSFRTRSHYAAEWDADFNRNGIIVIDEIPANGILNIALMGVSESGAPOTFTPDAL	420	
DB	393	KOHININAVRTSHYNDSDVWYELCNEYGLYVIDE-----NLETHG-----TWT-----	435	
QY	421	NDKQEAHKQAI-----RELARDKNHSAVVMMSIANEPASHEDGAREYF	465	
DB	436	--YIQEBOQKAVPSKPEKENVLDRCRSMYERKNNPSIIWLSGNSFGGEN-----F	488	
QY	466	EPLTNLTQLDPTRPITFANVGATYQLDRIISDLFDVSCINRYRWGVSQGDLEEAAL	525	
DB	489	QHMVTFEKEKSDTLVHYEGI--FHHRD-----YDASDIESTN--YVKPADVEERYALNN	538	
QY	526	EKEIHGMEKHFRIWTEYG--ADTLAHLISILGLPWSHEFQVQMLDMYHVRFDRIES	582	
DB	539	PK-----KPYLICEYSHMANGSGNLK-----YMWELFDQYPI	571	
QY	583	MAGEHVNMFAD-----FQTNLGIIRVDGN--KKGVFTRD--RKPKAAS	623	
DB	572	LOGGFIWMKQOALQATNEDGTSYLAGDGFDPNDGNFCGNGILIPADGTASFKIAEVK	631	
QY	624	---LRARTSID	632	
DB	632	KCYCPVKMTAVD	643	

RESULT 15
C83990
beta-galactosidase BH2723 [Imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004
C:Accession: C83990
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: AB8650; MUID:20512582; PMID:11058132
A:Accession: C83990
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-1014 <STO>
A:Cross-references: UNIPROT:Q9K9C6; GB:AP001516; GB:BA000004; NID:G10175192; PIDN:BA0644
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH2723
C:Superfamily: beta-galactosidase

Query Match	11.4%;	Score 382.5;	DB 2;	Length 1014;
Best Local Similarity	21.9%;	Pred. No. 2.6e-18;		
Matches 144;	Conservative 97;	Mismatches 253;	Indels 163;	Gaps 25;
QY	52	NLDGLMKFALASGLNDTAQPTAP-----LPKGLGECVPASYNDIFISREIH-----DHV	101	
DB	47	SLNGLMKFHYA-----INPWRPKREFYQGFDC--KCWDDILVPGHIOLOGYKQYV	97	
QY	102	GMVY-----YREVIYVKGWSGERLYVRAASATHHGRILYNNR	139	
DB	98	NWYFPMGDHHLRPPREIPEDDNPVGSYKCFDIPNNNSNHPLEISFGVETAFYVWLNGE	157	
QY	140	LVAEHVGCTPEEADVTLELVADGEKRLTIGVNNELTHETIPPGKITTGATGKRIQTYQ	199	
DB	158	FVGSSEDSFTPAEFDITSLVAVEG-----NKLCEYVQ-----RSTGSWLE--D	199	
QY	200	HDFVYVAGLARSIMLYSVPOOHIDITVTVDVG--DNGILNVEEVANQTTGOIQ--S	255	
DB	200	QDFWRMSGIFRDVYLYTIPNIVHVDHVRADLDRSLQGITLLEIKRSQEKEMIVAE	259	
QY	256	VIDEDGALVA-----KASAGQGYT--TIPSVKLMGCAAVLYQLOVNIIVGSSGVDVDTYNL	309	
DB	260	LYDAGAVAVATADMTNDQATVSNVS PALWSAEDPYLYKLLKLFDENGTIVEVVPQ	319	
QY	310	ATGVRTVTVKVASGQFLNGKPYFTFGKHEDTAVRGKGHDPAYVHDPQLMKWIGANSFR	369	
DB	320	KIGFRFELVNNITMLNGKRIYFKGVNRHEFNRGRTGVVTKEDMLDIKTKKKNINAVR	379	
QY	370	TSHYVAAEWDADFNRNGIIVIDEIPANGILNIALMGVSESGA-----PQTFPPDALN	421	
DB	380	TSHYPNNSWYQLOCEYGLYVIDE-----NLETHGSWOKGLVSESWNIPGNHL	429	
QY	422	DKQEAHKQAIRELARDKNHSAVVMMSIANEPASHED--GAREYFELTNLTQLDPTR	479	
DB	430	EWEPVIMDRAV--SMEERDNHPSIILWSCGNSYAGEYILNVSRYF-----KSVDSR	481	
QY	480	PITFANVGATYQLDRIISDLFDVSCINRYRWGVSQGDLEEAALAEKELHGM--OEKHF	538	
DB	482	LVHYEGVFARA-----YDATSDMESRMVAKPKDIEDYLTNDPRK	521	
QY	539	PIVMTY-----GADTLAHLISILGLPWSHEFQVQMLDMYHVRFDRIESMAGEHVNMFADPQ	595	
DB	522	PYISCEYMHMANGSLGGMKYTELE-----OKYPMYO-----GCFIMYID--	562	
QY	596	TNIGIIRVD-----GNK-----KGVFTDRKPKAAASLRARWTSI	631	
DB	563	--QALLKQRYGKEYFAYGDFGDRPTDYSFGANGIYVADRKPSPKMGVVKLYONI	617	

Search completed: March 18, 2005, 23:53:04
Job time : 46 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2005, 23:42:09 ; Search time 60 Seconds
(without alignments)
3492.718 Million cell updates/sec

Title: US-10-757-093-4
Perfect score: 3354
Sequence: 1 MKEFLGSLSLAPSLGTP.....RKPKAAHSLRARTSIDKN 634

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1401741 seqs, 330541175 residues
Total number of hits satisfying chosen parameters: 1401741

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/prodata/2/pubppa/US07_PUBCOMB.pep:*
2: /cgn2_6/prodata/2/pubppa/US06_NEW_PUB.pep:*
3: /cgn2_6/prodata/2/pubppa/US06_PUBCOMB.pep:*
4: /cgn2_6/prodata/2/pubppa/US07_NEW_PUB.pep:*
5: /cgn2_6/prodata/2/pubppa/US09_PUBCOMB.pep:*
6: /cgn2_6/prodata/2/pubppa/US08_NEW_PUB.pep:*
7: /cgn2_6/prodata/2/pubppa/US08_PUBCOMB.pep:*
8: /cgn2_6/prodata/2/pubppa/US09_PUBCOMB.pep:*
9: /cgn2_6/prodata/2/pubppa/US09_PUBCOMB.pep:*
10: /cgn2_6/prodata/2/pubppa/US09_PUBCOMB.pep:*
11: /cgn2_6/prodata/2/pubppa/US09_PUBCOMB.pep:*
12: /cgn2_6/prodata/2/pubppa/US09_PUBCOMB.pep:*
13: /cgn2_6/prodata/2/pubppa/US10_PUBCOMB.pep:*
14: /cgn2_6/prodata/2/pubppa/US10_PUBCOMB.pep:*
15: /cgn2_6/prodata/2/pubppa/US10_PUBCOMB.pep:*
16: /cgn2_6/prodata/2/pubppa/US10_PUBCOMB.pep:*
17: /cgn2_6/prodata/2/pubppa/US10_NEW_PUB.pep:*
18: /cgn2_6/prodata/2/pubppa/US11_NEW_PUB.pep:*
19: /cgn2_6/prodata/2/pubppa/US60_NEW_PUB.pep:*
20: /cgn2_6/prodata/2/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1714	51.1	618	14	US-10-356-088-28
2	1714	51.1	618	16	US-10-799-326-28
3	1714	51.1	1010	6	US-09-118-276-12
4	1714	51.1	1010	16	US-10-705-197A-12
5	1711	51.0	603	14	US-10-161-403-106
6	1711	51.0	603	14	US-10-120-145-6
7	1711	51.0	603	15	US-10-364-649-17
8	1711	51.0	603	15	US-10-364-649-23
9	1711	51.0	603	15	US-10-369-493-842
10	1706	50.9	604	10	US-09-893-525-37
11	1706	50.9	659	10	US-09-893-525-40
12	1706	50.9	850	10	US-09-893-525-42
13	1665.5	49.7	607	15	US-10-369-493-15291

14	1665	49.6	711	14	US-10-338-411-23	Sequence 23, Appl
15	1665	49.6	711	15	US-10-389-640-23	Sequence 23, Appl
16	1664.5	49.6	602	14	US-10-195-518-5	Sequence 5, Appl
17	1439	42.9	607	14	US-10-120-145-8	Sequence 8, Appl
18	1439	42.9	607	15	US-10-364-649-18	Sequence 18, Appl
19	1439	42.9	615	15	US-10-364-649-28	Sequence 28, Appl
20	1439	42.9	618	14	US-10-120-145-4	Sequence 4, Appl
21	1437	42.8	602	14	US-10-120-145-2	Sequence 2, Appl
22	1437	42.8	602	15	US-10-364-649-2	Sequence 2, Appl
23	1437	42.8	602	15	US-10-364-649-8	Sequence 8, Appl
24	1437	42.8	602	15	US-10-364-649-15	Sequence 15, Appl
25	1338	39.9	613	14	US-10-120-145-5	Sequence 5, Appl
26	1338	39.9	613	15	US-10-364-649-16	Sequence 16, Appl
27	1338	39.9	651	15	US-10-421-175-2	Sequence 2, Appl
28	1338	39.9	722	14	US-10-136-841-6	Sequence 6, Appl
29	1338	39.9	722	15	US-10-272-531A-6	Sequence 6, Appl
30	1338	39.9	722	15	US-10-272-483A-6	Sequence 6, Appl
31	1223	36.5	450	14	US-10-421-175-4	Sequence 4, Appl
32	1223	36.5	450	14	US-10-356-088-40	Sequence 40, Appl
33	1223	36.5	450	16	US-10-799-326-40	Sequence 40, Appl
34	1077.5	32.1	598	10	US-09-862-660-2	Sequence 2, Appl
35	1077.5	32.1	598	15	US-10-673-935-2	Sequence 2, Appl
36	993.5	29.6	376	15	US-10-364-649-4	Sequence 4, Appl
37	993.5	29.6	376	15	US-10-364-649-19	Sequence 19, Appl
38	988	29.5	372	15	US-10-364-649-3	Sequence 3, Appl
39	988	29.5	372	15	US-10-364-649-22	Sequence 22, Appl
40	883	26.3	563	15	US-10-364-649-6	Sequence 6, Appl
41	883	26.3	563	15	US-10-364-649-21	Sequence 21, Appl
42	882	26.3	563	15	US-10-369-493-2995	Sequence 2995, Ap
43	756.5	22.6	540	15	US-10-364-649-5	Sequence 5, Appl
44	741.5	22.1	408	15	US-10-369-493-13311	Sequence 13311, A
45	738	22.0	535	15	US-10-364-649-20	Sequence 20, Appl

ALIGNMENTS

RESULT 1
US-10-356-088-28
Sequence 28, Application US/10356088
Publication No. US20030167533A1
GENERAL INFORMATION:
APPLICANT: Yadvav, Narendra S.
TITLE OF INVENTION: Incein-Mediated Protein Splicing
FILE REFERENCES: C11806
US NA
CURRENT APPLICATION NUMBER: US/10/356,088
CURRENT FILING DATE: 2003-01-31
PRIOR APPLICATION NUMBER: US 60/354395
PRIOR FILING DATE: 2002-02-04
NUMBER OF SEQ ID NOS: 66
SOFTWARE: PatentIn version 3.1
SEQ ID NO 28
LENGTH: 618
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Modified GUS protein, with 6x His tags on the N-terminus and
OTHER INFORMATION: C-terminus
FEATURE:
NAME/KEY: PEPTIDE
LOCATION: (1)..(618)
OTHER INFORMATION: Modified GUS protein, with 6x His tags on the N-terminus and
OTHER INFORMATION: C-terminus
US-10-356-088-28

Query Match 51.1% Score 1714; DB 14; Length 618;
Best Local Similarity 55.2% Pred. No. 1.2e-136;
Matches 331; Conservative 92; Mismatches 161; Indels 16; Gaps 9;
40 VREPORTSRELVDIGLWKFAL--ASGLNDTAQPTAPLPKGLCPVPSYNDIFISRE 96
||| |::| ||||| |::| : : : : : ||| |

```

Db 9 VRPEVETPREIKKLDGLMAFSLDRENGCIDQ--RMMESALQESRALAVGSPFNDQFADAD 66
QY 97 IHHDVGVVYQREVIYPKGMSOERYLVRASATTHGRIVYNNRLVAEHVGGYTPFEADYT 156
Db 67 IRNVAGVWVYQREVEFIPKMGAGRIYLRFDATVHYGVWVNNQOEVEHOGGYTPFEADYT 126
QY 157 ELVAPGEKRLITGVNNELTHETIPPGKITTGNAATGRIQTYQHPFNVAAGLARSITWLS 216
Db 127 PYVIAKSVRIYVCNNELMWQITIPPGWIT--DENGKKQSYFHPFNVAAGLARSITWLT 185
QY 217 VPQOHTODITVTVDV--DGDNGILNVEVEVANQTTGQIQISVIDEDGALVAKASAGQTV 274
Db 186 TPTWTVDDITVTVTHVADCNHASVDWQV--VAN--GDVVELRDADQOVVATGQTSGLT 241
QY 275 TIPSVKLWOPGAAYLYOLOVNIYSSGDVDTYNLATGVTVKVASGQFLNKGPEYFTG 334
Db 242 QVNPPLMOPGEGYLYELCV--TAKSQTECDIYPLRGIRSAVVKGEQFLNKHKPEYFTG 299
QY 335 FGKHEDTAVRGKGDPAVYVHDFOLMKWIGANSFRTSHYPYAEVMDPADRNGIYVIDET 394
Db 300 FGHEDADLRGKGFVDVWVHDLMDWIGANSYRTSHYPYAEVMDPADRNGIYVIDET 359
QY 395 PAVGNIAL-MGVSESGAP-QTTPDAINDKTOEAHKQAIRBLIARDKXHASVVMMSIAN 452
Db 360 AAVGFPLSLGIGFEAGNKKPELYSEAVNGETOQAHLOAIKELIARDKXHASVVMMSIAN 419
QY 453 EPASHEDGAREYEPITNLTRQOLDPTRPITFANVGTATYQDRISLDFVSCINRYFGWY 512
Db 420 EPTROGAREYEPITNLTRQOLDPTRPITFANVGTATYQDRISLDFVSCINRYFGWY 479
QY 513 SQTGDEEAEALKEKELHGWQEKFRPIVMTYEGADTLAAGLSIILGPWSEEFQVQMLDM 572
Db 480 VQSGDLEAKVLEKELHGWQEKFRPIVMTYEGADTLAAGLSIILGPWSEEFQVQMLDM 539
QY 573 YHRVFDRIEMAGHEVWNPADFQTNIGIRVDGNKKGVFTDRPKRPAASHLPAWTSID 632
Db 540 YHRVFDRIEMAGHEVWNPADFQTNIGIRVDGNKKGVFTDRPKRPAASHLPAWTSID 599

RESULT 2
US-10-799-326-28
; Sequence 28, Application US/10799326
; Publication No. US2004012688A1
; GENERAL INFORMATION:
; APPLICANT: E.I. DuPont de Nemours, & Company
; APPLICANT: Yadav, Narendra S.
; APPLICANT: Yang, Jianjun Gene
; TITLE OF INVENTION: Intein-Mediated Protein Splicing
; FILE REFERENCE: CL1806 US CIP
; CURRENT APPLICATION NUMBER: US/10/799,326
; CURRENT FILING DATE: 2004-03-12
; PRIOR APPLICATION NUMBER: US 60/354395
; PRIOR FILING DATE: 2002-02-04
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 28
; LENGTH: 618
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Modified GUS protein, with 6x His tags on the N-terminus and
; OTHER INFORMATION: C-terminus
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(618)
; OTHER INFORMATION: Modified GUS protein, with 6x His tags on the N-terminus and
; OTHER INFORMATION: C-terminus
US-10-799-326-28

Query Match 51.1%; Score 1714; DB 16; Length 618;
Best Local Similarity 55.2%; Pred. No. 1,2e-136;
Matches 331; Conservative 92; Mismatches 161; Indels 16; Gaps 9;

```

```

QY 40 VRPQRTSRELVNLDGLMKFAL---ASGLNDTAPQWTAFLPKGLCECPVSPAYNDIFSRH 96
Db 9 VRPEVETPREIKKLDGLMAFSLDRENGCIDQ--RMMESALQESRALAVGSPFNDQFADAD 66
QY 97 IHHDVGVVYQREVIYPKGMSOERYLVRASATTHGRIVYNNRLVAEHVGGYTPFEADYT 156
Db 67 IRNVAGVWVYQREVEFIPKMGAGRIYLRFDATVHYGVWVNNQOEVEHOGGYTPFEADYT 126
QY 157 ELVAPGEKRLITGVNNELTHETIPPGKITTGNAATGRIQTYQHPFNVAAGLARSITWLS 216
Db 127 PYVIAKSVRIYVCNNELMWQITIPPGWIT--DENGKKQSYFHPFNVAAGLARSITWLT 185
QY 217 VPQOHTODITVTVDV--DGDNGILNVEVEVANQTTGQIQISVIDEDGALVAKASAGQTV 274
Db 186 TPTWTVDDITVTVTHVADCNHASVDWQV--VAN--GDVVELRDADQOVVATGQTSGLT 241
QY 275 TIPSVKLWOPGAAYLYOLOVNIYSSGDVDTYNLATGVTVKVASGQFLNKGPEYFTG 334
Db 242 QVNPPLMOPGEGYLYELCV--TAKSQTECDIYPLRGIRSAVVKGEQFLNKHKPEYFTG 299
QY 335 FGKHEDTAVRGKGDPAVYVHDFOLMKWIGANSFRTSHYPYAEVMDPADRNGIYVIDET 394
Db 300 FGHEDADLRGKGFVDVWVHDLMDWIGANSYRTSHYPYAEVMDPADRNGIYVIDET 359
QY 395 PAVGNIAL-MGVSESGAP-QTTPDAINDKTOEAHKQAIRBLIARDKXHASVVMMSIAN 452
Db 360 AAVGFPLSLGIGFEAGNKKPELYSEAVNGETOQAHLOAIKELIARDKXHASVVMMSIAN 419
QY 453 EPASHEDGAREYEPITNLTRQOLDPTRPITFANVGTATYQDRISLDFVSCINRYFGWY 512
Db 420 EPTROGAREYEPITNLTRQOLDPTRPITFANVGTATYQDRISLDFVSCINRYFGWY 479
QY 513 SQTGDEEAEALKEKELHGWQEKFRPIVMTYEGADTLAAGLSIILGPWSEEFQVQMLDM 572
Db 480 VQSGDLEAKVLEKELHGWQEKFRPIVMTYEGADTLAAGLSIILGPWSEEFQVQMLDM 539
QY 573 YHRVFDRIEMAGHEVWNPADFQTNIGIRVDGNKKGVFTDRPKRPAASHLPAWTSID 632
Db 540 YHRVFDRIEMAGHEVWNPADFQTNIGIRVDGNKKGVFTDRPKRPAASHLPAWTSID 599

RESULT 3
US-09-118-276-12
; Sequence 12, Application US/09118276
; Patent No. US20010011381A1
; GENERAL INFORMATION:
; APPLICANT: BABYCHUK, EBENA;
; APPLICANT: KUSHNIR, SERGEI;
; APPLICANT: DE BLOCK, MARC;
; APPLICANT: INZE, DIRK
; TITLE OF INVENTION: METHODS AND MEANS TO MODULATE PROGRAMMED
; TITLE OF INVENTION: CELL DEATH IN EUKARYOTIC CELLS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SIXBEY, FRIEDMAN, LEEDOM, & PERGUSON
; STREET: 8180 GREENSBORO DRIVE, SUITE 800
; CITY: MCLEAN,
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3-1/2" DISKETTE
; COMPUTER: IBM-COMPATIBLE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/118,276
; FILING DATE: 17-JUL-1998
; PRIOR APPLICATION DATA: NONE
; ATTORNEY/AGENT INFORMATION:
; NAME: SCHULMAN, ROBERT M.; SALEN, KENNETH H.
; REGISTRATION NUMBER: 31,196; 43,077
; REFERENCE/DOCKET NUMBER: 6201-0003

```


TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 790-9110
 TELEFAX: (703) 883-0370
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1010 RESIDUES
 TYPE: AMINO ACID
 STRANDEDNESS: SINGLE
 TOPOLOGY: LINEAR
 US-09-118-276-12

Query Match 51.1%; Score 1714; DB 9; Length 1010;
 Best Local Similarity 55.2%; Pred. No. 2,5e-136;
 Matches 331; Conservative 92; Mismatches 161; Indels 16; Gaps 9;

```

QY 40 VRPQRTSSRELVDLMKFL--ASGLNDTPQPTAPLPGKECPVPSYNDITSR 96
DB 409 VRPEPTREIKKLDLMFSLDRENGCIDQ--RWMESALQESRAIAPGSEFNDQPAD 466
QY 97 IDHVGWVYVQREVIYPKGMSOERYLVRAESATTHGRIVNNRLVAEHVGGYTPFEADYT 156
DB 467 IRNYAGNVWYQREVFIPKMGAGRIYLRPAVTHYGVKWNNOEVEHGGYTPFEADYT 526
QY 157 ELVAPGEKRLITGVNNELTHETIPPGKITTGNAKRIQTYOHDFYNYAGLARSIMLYS 216
DB 527 PYVIAGKSRIYVCNNELMQITIPPGMVI--DENGKKQSYFHDFFNYAGIHRSMVLYT 585
QY 217 VPQOHIQDITVTVDV--DGDNGILNVEVANOQTGOIQISVIDEDGAIYAKASGAQGY 274
DB 586 TPRTWVDITVTYVHVAQDCNHAQSVDMQV--VAN--GDVSEVLADADQOVVATGQTSGLT 641
QY 275 TIPSVLNPGGAAYLYOLOVNIYVSSGDVVDYTNLATGRTYVYVAGSOPFLINKPPEYFTG 334
DB 642 QVNVPHLMQPGSEYVLELCV--TAKSQTECDIYPLRVGIRSAVAVKGEQFLINKPPEYFTG 699
QY 335 FGKHEDTAVRGKGDPAVYVHDFOLMKWIGANSFRTSHYPYAEVWDFADRNQIYVIDET 394
DB 700 FGRHEDADRGKGFNDVNLVHDAIMDWIGANSYRISHYPYAEMLDMDAHEGIVVIDET 759
QY 395 PAVGNIAL--MGVSESGAP--QTFPDALNDKTOEAKKAIKRELIADKXHASVVMMSIAN 452
DB 760 AAVGFNLISGIGPEAKNKPELYSEAVNGETOQAHLQAIKELIADKXHASVVMMSIAN 819
QY 453 EPASHEGAREVEPEPTNLTRQDPTPTTFANVGATATYQDRISLDFVSCINRYFGWY 512
DB 820 EPDTRQAGREVPAPLAETRKLDPTFRPITCVNVMFCDAHTDITSDLPVLCINRYFGWY 879
QY 513 SQTGLEEAALAEKELHGOEKFRHPIWTEYGADTLGLSHILGLPMSSEFOVQMLDM 572
DB 880 VQSGDLETAKEVLEKELHMOEKHQPITITEYGVDTLGLSHMTYDMSEETQCAMLDW 939
QY 573 YHRVPRISMAGEHYVNFADPOTNLGIIRVDGNKKGVFTDRKKPYAAHSLPARWTSID 632
DB 940 YHRVFPDRISAVVGEQVWNFADPATSQILRVGNNKKGIPTDRKKPKSAALFLQKRWTKWN 999

```

RESULT 4
 US-10-705-197A-12
 ; Sequence 12, Application US/10705197A
 ; Publication No. US20040128704A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Elena, Babyzhuk
 ; APPLICANT: Sergei, Kuznet
 ; APPLICANT: Marc, De Block
 ; TITLE OF INVENTION: Methods and means to modulate programmed cell death in eukaryotic
 ; FILE REFERENCE: 58764.0000039
 ; CURRENT APPLICATION NUMBER: US/10/705,197A
 ; PRIOR FILING DATE: 2003-11-12
 ; PRIOR APPLICATION NUMBER: US 09/118,276
 ; PRIOR FILING DATE: 1998-07-11
 ; NUMBER OF SEQ ID NOS: 21
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 12

LENGTH: 1010
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: fusion protein between APP N-terminal domain and GUS protein
 US-10-705-197A-12

Query Match 51.1%; Score 1714; DB 16; Length 1010;
 Best Local Similarity 55.2%; Pred. No. 2,5e-136;
 Matches 331; Conservative 92; Mismatches 161; Indels 16; Gaps 9;

```

QY 40 VRPQRTSSRELVDLMKFL--ASGLNDTPQPTAPLPGKECPVPSYNDITSR 96
DB 409 VRPEPTREIKKLDLMFSLDRENGCIDQ--RWMESALQESRAIAPGSEFNDQPAD 466
QY 97 IDHVGWVYVQREVIYPKGMSOERYLVRAESATTHGRIVNNRLVAEHVGGYTPFEADYT 156
DB 467 IRNYAGNVWYQREVFIPKMGAGRIYLRPAVTHYGVKWNNOEVEHGGYTPFEADYT 526
QY 157 ELVAPGEKRLITGVNNELTHETIPPGKITTGNAKRIQTYOHDFYNYAGLARSIMLYS 216
DB 527 PYVIAGKSRIYVCNNELMQITIPPGMVI--DENGKKQSYFHDFFNYAGIHRSMVLYT 585
QY 217 VPQOHIQDITVTVDV--DGDNGILNVEVANOQTGOIQISVIDEDGAIYAKASGAQGY 274
DB 586 TPRTWVDITVTYVHVAQDCNHAQSVDMQV--VAN--GDVSEVLADADQOVVATGQTSGLT 641
QY 275 TIPSVLNPGGAAYLYOLOVNIYVSSGDVVDYTNLATGRTYVYVAGSOPFLINKPPEYFTG 334
DB 642 QVNVPHLMQPGSEYVLELCV--TAKSQTECDIYPLRVGIRSAVAVKGEQFLINKPPEYFTG 699
QY 335 FGKHEDTAVRGKGDPAVYVHDFOLMKWIGANSFRTSHYPYAEVWDFADRNQIYVIDET 394
DB 700 FGRHEDADRGKGFNDVNLVHDAIMDWIGANSYRISHYPYAEMLDMDAHEGIVVIDET 759
QY 395 PAVGNIAL--MGVSESGAP--QTFPDALNDKTOEAKKAIKRELIADKXHASVVMMSIAN 452
DB 760 AAVGFNLISGIGPEAKNKPELYSEAVNGETOQAHLQAIKELIADKXHASVVMMSIAN 819
QY 453 EPASHEGAREVEPEPTNLTRQDPTPTTFANVGATATYQDRISLDFVSCINRYFGWY 512
DB 820 EPDTRQAGREVPAPLAETRKLDPTFRPITCVNVMFCDAHTDITSDLPVLCINRYFGWY 879
QY 513 SQTGLEEAALAEKELHGOEKFRHPIWTEYGADTLGLSHILGLPMSSEFOVQMLDM 572
DB 880 VQSGDLETAKEVLEKELHMOEKHQPITITEYGVDTLGLSHMTYDMSEETQCAMLDW 939
QY 573 YHRVPRISMAGEHYVNFADPOTNLGIIRVDGNKKGVFTDRKKPYAAHSLPARWTSID 632
DB 940 YHRVFPDRISAVVGEQVWNFADPATSQILRVGNNKKGIPTDRKKPKSAALFLQKRWTKWN 999

```

RESULT 5
 US-10-161-403-106
 ; Sequence 106, Application US/10161403
 ; Publication No. US20030119104A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Perkins, Edward
 ; APPLICANT: Perez, Carl
 ; APPLICANT: Lindbaum, Michael
 ; APPLICANT: Greene, Amy
 ; APPLICANT: Leung, Josephine
 ; APPLICANT: Fleming, Elena
 ; APPLICANT: Stewart, Sandra
 ; APPLICANT: Shellard, Joan
 ; TITLE OF INVENTION: CHROMOSOME-BASED PLATFORMS
 ; FILE REFERENCE: 24601-420
 ; CURRENT APPLICATION NUMBER: US/10/161,403
 ; PRIOR FILING DATE: 2002-05-30
 ; PRIOR APPLICATION NUMBER: 60/294,758
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: 60/366,891
 ; PRIOR FILING DATE: 2002-03-21

ORGANISM: *Escherichia coli*
US-10-364-649-17

Query March	51.0%;	Score 1711;	DB 15;	Length 603;
Best Local Similarity	55.0%;	Pred. No. 2e-136;		
Matches 330;	Conservative	93;	Mismatches 161;	Indels 16;
				Gaps 9;

Qy	40	VRPGQTSRELLVNLGKMFAL---ASGLNDTAQOPTAPLPKLECECPVPSYNDIFLSRE	96
Db	2	LRPVEETPRREIKKLQGLNAFSLDREKCGIDQ--RWMESHQESRALAIVGSPFDQPADAD	59
Qy	97	IDHWGVYYQREIVIPKQMSQERYLVRASATHHGRIVYNNRLVAEHVGYTPFEADVT	156
Db	60	IRNVAQNVWYQGEVPIPKMGAGQRIYLRFDAVTHYGVKAWNNQGEVMEHQGYTPFEADVT	119
Qy	157	ELVABGEKRLTGVNNELTHETITPPGKITTGATGKRLQTYOHDYVNTAGLARSLTWLS	216
Db	120	PYVJAGKSVRIIVCVNNELMWQTIPEGWYIT--DENKKQOSYPHDEFENYAGHRSWLVT	178
Qy	217	VPGQHOIOITVTVTDV--DGDNGLINVEVNAVNOTTGQIOISVDEGAIVAKASGAQGTV	274
Db	179	TPNTWVDITVTVTHVAQDCNHAASVDQV--VAN--GDVSVELADADQOVATGQTSGLT	234
Qy	275	TIPSVKLMQPGAAYLYOLQVNIIVGSSGDVVDTYNLATGVRTVYKASQOPFLNKPPEYFTG	334
Db	235	QVNNVHLMQPGGXYIELCV--TASQTECDIYPLAVGIRSAVAKGEQFLINKKPFYFTG	292
Qy	335	FGKHEDTAVRGKGHDPYAVVHDFOLMKWIGANSFRTSHYPRYAEVMDPADRNGIIVIDET	394
Db	293	FGRHEDADLRGSGFNPVLMVHDHALMDWIGANSYRSHSPRYAEEMDMADBEHGIIVIDET	352
Qy	395	PVAGVNLAL--MOWSSGAP--QTETPDAINDKQOEAKOIRELIARDKNHASVVMSSIAN	452
Db	353	AAVGNNLSLGIFEPAGNKPKXEYSEBAVNGEFTQALQAKELIARDKNHPSVVMSSIAN	412
Qy	453	EPASHEDGARREYFEEPLTNLTROLDPRPTTFANVGATATQOLDRIIDLFDVSCINRRFYGWY	512
Db	413	EPDRPRQGARREYFALAEKTRKLDPRPTTCVNVMECDANTDITISDLFLCINRRYGYWY	472
Qy	513	SGTGDLLEAEALAEKLELHGQEKFRHPIVWTEYAGDTLAGLHSILGLPMSEEFQOVMLDM	572
Db	473	VQSGGLETAEBKYLEKEILAMQEKLHPITITEYGVDVTLGLHSMYTDMSSEBYQCAWLDM	532
Qy	573	YHRVDRDRLESMAGEHVWNPADPOTNIGIRVNDGNKGVFTRPDKKPKRAAHSLPARATSID	632
Db	533	YHRVDFRVSAAVGEQVWNPADPATSQGLIRVGNKKGIFTRDKKPKSAALFLQKRWGTGN	592
RESULT 8			
US-10-364-649-23			
Sequence 23, Application US/10364649			
Publication No. US20030229921A1			
GENERAL INFORMATION:			
APPLICANT: Richard A. Jefferson and Jorge E. Mayer			
TITLE OF INVENTION: MICROBIAL B-GLUCURONIDASE GENES, GENE			
TITLE OF INVENTION: PRODUCTS, AND USES THEREOF			
FILE REFERENCE: 190106.405C1			
CURRENT APPLICATION NUMBER: US/10/364,649			
CURRENT FILING DATE: 2003-02-12			
PRIOR APPLICATION NUMBER: 10/364,649			
PRIOR FILING DATE: 2003-02-12			
PRIOR APPLICATION NUMBER: US 09/270,957			
PRIOR FILING DATE: 1999-03-17			
NUMBER OF SEQ ID NOS: 112			
SOFTWARE: FastSeq for Windows Version 4.0			
SEQ ID NO 23			
LENGTH: 603			
TYPE: PRT			
ORGANISM: Escherichia coli			
US-10-364-649-23			

Matches 330; Conservative 93; Mismatches 161; Indels 16; Gaps 9;

```

QY      40 VRPQRSSSELVLNLDLMLKFAI---ASGLINDPAQWPTAPLPGLGECPPASPANDTFFISRE   96
Dd      2 LRREVFETPEIKELDLMAFSLDRENGCGIDQ--RMWESALQDSRAIAPGSDNDOPADAD    59
QY      97 IHDFVGVVVYOREVIVIPKWSOERYLVRAESAETHHGRITYNNRLVAEHVGTYTPREADVT   156
Dd      157 ELVAPBEKRLLITGVNNNELTHETIPPGKITTGNAIGKRIQTQHDPFYVAGLAGARIMLYS   216
QY      120 PYVIACKSVRITVCVNNNELMQTIPEGWAIT-DENGGKKOSSYFHFPFYAGIHRSVMILYT   178
Dd      217 VPQOHODITVTVTDV--DDNGLINVEYEVANQTTGOLOISVTEBDAIIVAASAOGCTV   274
Dd      179 TPTWTWDDITVTVTHAQDCNHASVDNQ-VAN--GDVSVELRDADDOOVATGCSTSGTL   234
QY      275 TITSVYLMOPGAAYLYQLQVNIIVSGSSGDVDTYNLATGVRTVKVASOFLINGKBPFTYG   334
Dd      235 QVNVPHLMQGBEYLIELCV--TAKSQTECDIYPLRVGRSAVAVAGEOFLINHKKPFYFTG   292
QY      335 FGHGEDTAVRKGKHDEAVYVHADFOLMKIIGANSFRTSHYPVAEEVMDPADRNGIYVIDET   394
Dd      293 FGHEHDADRKGCFDNVLVWHDLALMDWIGANSYRTSHYPVAEMLDMWADEHGIVIDET   352
QY      395 PANGIALAI-MGVSEEGAP-QTFTPDAINDKTQEAHKAIRLIARDKNHASVVMWMSIAN   452
Dd      353 AAGFGNLISGI GREAGNKCKPELYSEBAVGEVQAHLQAIKELIARDKNHPSEVWMSIAN   412
QY      453 EPVSHEDGAREREYEPILTNLTROLDPRPTTFPANVGATGYOLRISIDLFPVGSINRYFCGY   512
Dd      413 EPTRPQGAEREFAPLAETARKIDPRPTTCVNMFCAHTDDITSDLEPDVLCINRYICY   472
QY      513 SQTGDEEAAALEKEKLHMOKFKHRPIVMEYEGADTLAHLHSIIGLPWSEBFQVOMLDM   572
Dd      473 VQSGDELTEAEKYLEKELLAMQEKLNQPIITTEYGVDTLAGLSMTYTWMMSEBYQCAMLDM   532
QY      573 YHRVFDRIISMAGEHVNFADFOTNLGIITVDGNKKGVETRRDKRKPKAAASHRAPWTSID   632
Dd      533 YHRVFDRISSAVVGEQWNFADFATSGLILRVGNNKKGIPTRRDKRKPSAAFLIQKWTKGMN   592

RESULT 9
US-10-369-493-842
Sequence 842, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052) B
CURRENT APPLICATION NUMBER: US/10/369,493
PRIOR FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 842
LENGTH: 603
TYPE: PRP
ORGANISM: Escherichia coli
US-10-369-493-842
```

```

04 40 VRPORTSSRELVL DGLWKAL--ASGINDTAQPTAPLPKGIECPVPASINDIFISRE 96
    ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

```

Db	2	LRPVEPTREIKKLDGLMAFSLDRENCIGIDQ--RMWESALQESRALAIVAGSFNDQPADAD	55
Qy	97	IHDHGVNTYYQGEVIVLPKGMQSQRILYVRAESATHGRILYNNNLVAEHVGGTTPPEADVT	156
Db	60	IRNVAGNWWYQREVEFIPKMAQQRILRPDPAVHYKVMWNNQEVNEHQGGYTPPEADVT	119
Qy	157	ELVAAGEKRLTIGVNNELTHETIPPGKLTGNGATGKRIQIOTYOHDPYVAGLARSIMLYS	216
Db	120	PVYVAGKSVRIIVCVNNELMOTIIPGMVIT--DENKKKQOSYFHDFFNAGIHRSVMLYT	178
Qy	217	VPQGHIODITVNTDV--DGDNGLINYEVEVANQTTGQIQISVIDEDGAIYAKASGAQGYT	274
Db	179	TEPNTWDDITVTVTHAQCDCNHASVDMQV-VAN--GDVSEVLBDADQGVVATGQGTSGTL	224
Qy	275	TIPSVKLMQPGAAIYLQLOVNIYVSGSDVDVDTYNLATGYRTYKVASGQFLINCKPEFYFG	334
Db	235	QVNVPHLMQPGSGIYLCLCV--TAKSQTECDIYPLKVGIRSAVVKGEQFLINHKPEFYFG	292
Qy	335	FGKHEDTAVRGKGHPPAYWVHDFOLMKWIGANSFRTSHYPVAEEVMDPDRNGIIVIDET	394
Db	293	FGRHEDADLRGCGPNDVLMVHHDHALMDWIGANSYRSHIPYAEEMLDMDDEHGIIVIDET	352
Qy	395	PAVGINIAL-NGVSSGAP-QFTTPDAINDKTOEAHKQAIRBLIARDKXNASVVMSSIAN	452
Db	353	AAVGFNLSIGIGFEAGNKKPELYSEBAVNGETQAHIAIKELIARDKXNPSVVMSSIAN	412
Qy	453	EPASHEDGARVEFEELTNLTROLDPRTTFANVGNATYQDRIISLPFVSCINRFEGY	512
Db	413	EPDTRPQGRRETFAPLAETKRLDPRITTCVNVVFCDAHTTISLPFVCLINRYYGV	472
Qy	513	QQTGDLLEAEAALEKEHLHMOEKFHRPIYVTEYGDATLGLHSIIGLPMSEEFQVOMLDM	572
Db	473	VQSGDLETAEKVLEKELLAMQEKLMQPIITTEYGVDTLAGHSMTYDMSSSEYQCAMLDM	532
Qy	573	YHRVDRIRLESAGGEVWVNFADQTNLGIIRVUNGKKGVTRORRKPAAAHSLPARVTSID	632
Db	533	YHRVDFRVSAAVVGQVWVNFADPATSQGILIRVGNKKGIPTDRKPKPSAAFLQKRWTKGN	592
RESULT 10			
US-09-893-525-37			
Sequence 37, Application US/09893525			
Publication No. US20030126631A1			
GENERAL INFORMATION:			
APPLICANT: Moloney, Maurice M.			
APPLICANT: Van Rooijen, Gijb			
TITLE OF INVENTION: Preparation of Heterologous Proteins on Oil Bodies			
FILE REFERENCE: 9389-172			
CURRENT APPLICATION NUMBER: US/09/893,525			
CURRENT FILING DATE: 2001-06-29			
PRIOR APPLICATION NUMBER: US 09/210,843			
PRIOR FILING DATE: 1998-12-15			
PRIOR APPLICATION NUMBER: US 08/846,021			
PRIOR FILING DATE: 1997-04-25			
PRIOR APPLICATION NUMBER: US 08/366,783			
PRIOR FILING DATE: 1994-12-30			
PRIOR APPLICATION NUMBER: US 08/142,418			
PRIOR FILING DATE: 1993-11-16			
PRIOR APPLICATION NUMBER: US 07/659,835			
PRIOR FILING DATE: 1991-02-22			
NUMBER OF SEQ ID NOS: 42			
SOFTWARE: PatentIn version 3.1			
SEQ ID NO 37			
LENGTH: 604			
TYPE: PRT			
ORGANISM: Artificial Sequence			
FEATURE:			
OTHER INFORMATION: Phas-GUS-phaas			
US-09-893-525-37			

Query Match	50.9%;	Score 1706;	DB 10;	Length 604;
Best Local Similarity	54.8%;	Pred. No. 5.5e-136;		
Matches 329;	Conservative 94;	Mismatches 161;	Indels 16;	Gaps 9;

Qy	40	VRPQPSREELVNDLMEFEAL---ASGLNDPAQWPTLPKGLACCPAPAYNIJSRE	96
Db	3	LRVELEPTREIKKLDDGLMAFSLDRENCIGDQ--RMBSALQESRAIAPGSGFNQFADAD	60
Qy	97	IHDVGVVYVYQREVIIVPKMSQERYLVRAESATIHGRITVNNRLVAEHVGGYTFPEADVT	156
Db	61	IRVYAGVWVYQREVFIPKMGAGQRLVLRDAVTHGKVMVNNQGEVMEHGQYTFPEADVT	120
Qy	157	ELVAPEKRLRLITGVNNELTHETTPGKITTSNATGKRIQTYQNDFTVYAGLAISIMLYS	216
Db	121	PVYIAKSVRIITVCVNNELMTQITPEGWIT--DENGKKKQSYFHDFFNYAGIHSVMLYT	175
Qy	217	VPRQHIQDITTVWTV--DGDNGLIYEVENVNQTTGQIQIVIDEBAQIVKASGAGTV	276
Db	180	TEPTWDDITTVTHVQAQDNHASVMOY-VAN--GDVSVELRADDQGVATGQGTGTL	233
Qy	275	TIPSVKLMPQGAAYLYLOVNIIVGSSGVVDPTVNLATVTRTVKASGQFLNGKRFYFTG	334
Db	236	QVNVPHLMQPGSGYLVELCV--TAKSQRECDITPLRVGIRSVANAKGQGLNHNKRFYFTG	299
Qy	335	FGHEBDTAVRGKCHDPAVYVHDFOLMKNIIGANSFRTSHYPAAEVMYDADRNGIVIDET	394
Db	294	FGHEBDADRGKGFVTVLVVHDHALMDWIGANSYRTSHYPAAEMLDMDADHGIVIDET	353
Qy	395	PANGNLIAL-MGVSSSGAP-QTFETDADNDKTOEHKQATIELLARPKNASVVMWSIAN	455
Db	354	AAAGFSLSIGIGEAQNKPELSEAVANGFTQQAHAQAIKELLARPDNHSVVMWSIAN	413
Qy	453	EPASHEBDGAREFEPTVNLTRQLDPTRPITFPAVNTATATYQDRIISDLFDVSCIRRYGMY	512
Db	414	EPDTRQGARREVFAPLAELATKRLDPTRPITVCVNWPCDANHDTISDLFDVCLNRYGMY	473
Qy	513	SQTGDLEEAALBEKLEHGMQEKFRPIVMTYEGADTLAHLSTLGLPWSEEFQVQMLDM	572
Db	474	VQSGDLETEKYLEKELLAMQEKHQPIITIEYGVDTLAGHSYVTWMSSEYQCAWILDM	533
Qy	573	YHHPVPRISMGHEVNVMPDQTNLGLIRVDQKKGVPTFDRPKAAHSLRARWMSID	632
Db	534	YHHPVPRVAVVGEVNVMPADATISQGLIRVGANKKGITFDRPKSAALFLQRRWGMN	593

```

RESULT 11
US-09-893-525-40
Sequence 40, Application US/09893525
Publication No. US20030126631A1
GENERAL INFORMATION:
APPLICANT: Moloney, Maurice M.
APPLICANT: Van Rooijen, Gijf
TITLE OR INVENTION: Preparation of Heteroc
FILE REFERENCE: 9369-172
CURRENT APPLICATION NUMBER: US/09/893,525
CURRENT FILING DATE: 2001-06-29
PRIORITY APPLICATION NUMBER: US 09/210,843
PRIORITY FILING DATE: 1998-12-15
PRIORITY APPLICATION NUMBER: US 08/846,021
PRIORITY FILING DATE: 1997-04-25
PRIORITY APPLICATION NUMBER: US 08/366,783
PRIORITY FILING DATE: 1994-12-30
PRIORITY APPLICATION NUMBER: US 08/142,418
PRIORITY FILING DATE: 1993-11-16
PRIORITY APPLICATION NUMBER: US 07/659,835
PRIORITY FILING DATE: 1991-02-22
NUMBER OF SEQ ID NOS: 42
SOFTWARE: PatentIn version 3.1.
SEQ ID NO 40
LENGTH: 659
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
US-09-893-525-40
OTHER INFORMATION: phas-oleo GUS-phas

```

OTHER INFORMATION: phas-oleo GUS-phas
US-09-893-525-40

OTHER INFORMATION: phae-caleo-GUS-phae

US-09-893-525-42

Query Match 50.9%; Score 1706; DB 10; Length 850;
Best Local Similarity 54.8%; Pred. No. 9, 1e-136;
Matches 329; Conservative 94; Mismatches 161; Indels 16; Gaps 9;

QY 40 VRPQRTSSRELVNLDGIMKFL--ASGLNDTPQPTAPLPGLGSEPPVASTNDIFISRE 96
DB 249 LRPEETREIKKLIDGIMAFSLDRENGCIDQ--RWMSALQESRAIAYGSSFNDDPADAD 306
QY 97 IHHDVGVVQREVEIYPKGMSOERYVRAASATTHGRIVNNRLVAEHGVTTPREADVT 156
DB 307 IRNYAGVWVQREVFIPKMGAGQIVLRFDATYTHGKVVANNQSEVMEHOGGVTTPREADVT 366
QY 157 ELVAPGEKFLITGVNNELTHETIPGKLTGNGATGRIQTYOHDFFYNYGLARSIMYS 216
DB 367 PYVLAGSVARITVCANNELMTQITPPGWIT-DENGKKQGSFHPFPVYAGIHSVMLYT 425
QY 217 VPQOHHODITVTVDV--DGDNGLINEVEVANNQTTGQIQISVIDEGATVAKASGAQTV 274
DB 426 TPTWTVDDITVTVTHVADCCNHASVDQV-VAN--GDVSVELRDADQGVATGCGTGT 481
QY 275 TTSVTLMGGAAYVLOQVNIYVSSGDVVDVYNLATGRTKTVAGSOFELINGKPFYFG 334
DB 482 QVNPVPLMQEGEYLYELCV-TAKSQTECDIYLRVIGIRSVAVGQOCLINHKPFYFG 539
QY 335 FGKHEDTAVGKGDHDAVYVWDFOLMKWIGANSFRTSHPYAEVWDPADRNGIIVIDET 394
DB 540 FGRHEDADLRGKFPDVLVWHDALMDWIGANSYRSHPYAEMLDMADDEHGIIVIDET 599
QY 395 PAVGLNIAL-MGVSESGAP-QTFPTDAINDKTOEAAKQAIIRLIARDKXHASVVMWSTAN 452
DB 600 AAVGFSISLIGIEAGNKPKELYSEBAVNGETQOAHLOAIKELIARDKXHPSEVMMSTAN 659
QY 453 EPASHDEGAEEYEPPLTNLTROLDPRPLTPFANVGATVQLRISGLDPVGSINKYFPGY 512
DB 660 EPTRPGAEYEPAPLAEARKLDPRLPYCVNVWFCDAAHTTISLDFVLCINLYGY 719
QY 513 SQTGDEEAAALEKEKELHGWQEKHRRPIVWTEYGAQDTLAGLSIIGLPSESEFQVQMLDM 572
DB 720 VQSGDLETAKEVLEKELMLAQEKLNHPILITTEYGVDTLAGLSMTMDMSESEYQCAMIDM 779
QY 573 YHRVFDRIESMAGEHVNFADPQTNIGIIRVDGNKKGVETFRDKRPPKAAHSLPARWTSID 632
DB 780 YHRVFDRESVAVVEQVWNFADPATSQILRVGKNKKGIETFRDKRPPKSAFLLOKSWTGMN 839

RESULT 13
US-10-369-493-15291
Sequence 15291, Application US//10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 15291
LENGTH: 607
TYPE: PRP
ORGANISM: Agrobacterium tumefaciens
US-10-369-493-15291

Query Match 49.7%; Score 1665.5; DB 15; Length 607;

Best Local Similarity 53.8%; Pred. No. 1.5e-132;
Matches 324; Conservative 94; Mismatches 167; Indels 17; Gaps 10;

```

QY 38 IKVRPORTSSRELVLNDGLMKFAL--ASGLNDTAQPMVAPLPGKECVPPVSYNDIFLS 94
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 5 LMLRPVETPTREIKLDDGLMAFSLDRENGCIDQ--RMWESALQESRAIAVPSPFDQRAD 62
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 95 REIHDVGVWYQREYIVPKGMSQERYLVRAESATHGRIVYNNRLVAEHVGYTPPEAD 154
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 63 ADIRNAGVWYQREYIVPKGMAQGRIVRPAVTHYKVMWVNGQVMEHQGYTPPEAD 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 155 VTELVAPEKFRITIGNNELTHETTPPKITTGATGKRIQTYQDPEFNAGLRSIML 214
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 123 VTEYVLAGSKSVRITVCNNELMQITIPGKVTIT--DENGKKOSYFFDFNFVAGIHSVNL 181
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 215 YSPQOHIQDITVTVTV--DGDNGLINYEVEVANQTTGOIQLSVDEDAIVAKASGAG 272
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 182 YTPENTWVDITVTVTHVAQDCNHSYDMQV--VAN--GDVSVELRADQGVVATGQTS 237
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 273 TVTIPSVKMQPGAAYLYQLOVNIYVSSGDVDTYNLATGVTYKVASGOFLNGKPYF 332
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 238 TLQVNVPHLMQPGGYLYELCV--TAKSQTECDIYPLRVGIRSVAVKGOFLINHKPYF 295
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 333 TGEGKEDTVVRKGNHDPAYMVDHDPOLMKRIGANSFRTSHYPAAEYMPDADNGIYVID 392
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 296 TGFGRHEDADLRGKGFNVLMVDHMLMDVIGANSYTSYHYPAEEMLDMADEHGIYVID 355
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 393 ETTPAVGNIAL--MGVSSGAP--QTFETPAINDKTOEAKQAIRELJARDKNSAVVMWSI 450
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 356 ETAAVGNLSLIGIFPAKGNPKELYSBAVNGETQAHLOAIKELLARDQNHSSVVMWSI 415
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 451 ANEPASHEDGAREYFEPVLNLTROLDPTPTPTANVTATYQDRIISLFDVSCINRYFG 510
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 416 ANEDPTPQ--VHGNISPLAETRLDPTPTTCVNVFCAHTDTISDLPTVCLNRYG 474
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 511 MYSTGLLEEAALKEKLEHGMOKFRPIYMTVEYAGDTLAGLSILGLPMWSEFQVOML 570
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 475 WYVOSGDIETKEKLEKELAMOKELHQPIITTEYGVDTLAGHSVMTDMWSEYQCAWL 534
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 571 DMVHRVFDRIISMAGEHVMNFPQTNLGIIRVDGNKKGVFTDRKRKAANAHSIRAWTS 630
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 535 DMVHRVFDRIISMAGEHVMNFPQTNLGIIRVDGNKKGVFTDRKRKAANAHSIRAWTS 594
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 631 ID 632
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 595 MN 596
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 14
US-10-338-411-23
Sequence 23, Application US/10338411
Publication No. US20030153045A1

GENERAL INFORMATION:
APPLICANT: Butt, Tauseef
APPLICANT: Weeks, Stephen
APPLICANT: Tran, Hiep
APPLICANT: Malakhova, Oksana
APPLICANT: Malakhov, Micheal
TITLE OF INVENTION: Methods and Compositions for Protein
FILE REFERENCE: 1955-2792US1
CURRENT APPLICATION NUMBER: US/10/338,411
CURRENT FILING DATE: 2003-01-07
PRIOR APPLICATION NUMBER: 60/346,449
PRIOR FILING DATE: 2002-01-07
NUMBER OF SEQ ID NOS: 65
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 23
LENGTH: 711
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic Sequence

US-10-338-411-23

Query Match 49.6%; Score 1665; DB 14; Length 711;
Best Local Similarity 52.4%; Pred. No. 2.2e-132;
Matches 328; Conservative 90; Mismatches 178; Indels 22; Gaps 12;

```

QY 19 TPAARHFRPMNMD--HEQPL----IKVRPORTSSRELVLNDGLMKFAL--ASGLNDTAQ 70
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 85 TPELDMEENDITLAEHQIGCMEMFMRPVETPTREIKLDDGLMAFSLDRENGCIDQ--R 142
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 71 PWTAPLEKLECPVPASYNDFISREIHDVGVWYQREYIVPKGMSQERYLVRAESATH 130
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 143 WMSALQESRAIAVPSPFDQRADIRNAGVWYQREYIVPKGMAQGRIVRPAVTH 202
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 131 HGRIVNNRLVAEHVGYTPPEADVTELVAPEKFRITIGNNELTHETTPPKITTGNA 190
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 203 YGKVMWVNGQVMEHQGYTPPEADVTVYVLAGSKSVRITVCNNELMQITIPGKVTIT--DE 261
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 191 TGKRIQTYQDPEFNAGLRSIMLYSPQOHIQDITVTVTV--DGDNGLINYEVEVANQ 248
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 262 NGKKQSYFDFNFVAGIHSVNLVTPNTWVDITVTVTHVAQDCNHSYDMQV--VAN-- 318
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 249 TGQIOISVIDEDGAIYAKASGAGTPTIPSVKMQPGAAYLYQLOVNIYVSSGDVDTYN 308
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 319 -GDVSVELRADQGVVATGQTSGLQVNVPHLMQPGGYLYELCV--TAKSQTECDIY 375
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 309 LATGVTYKVASGOFLNGKPYFTEGKEDTVVRKGNHDPAYMVDHDPOLMKRIGANSF 368
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 376 LRVGIRSVAVKGOFLINHKPYFTGFGRHEDADLRGKGFNVLMVDHMLMDVIGANSY 435
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 369 RTSYHYPAEYMPDADNGIYVIDETPAVGNIAL--MGVSSGAP--QTFETPAINDKTOE 426
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 436 RTSYHYPAEYMPDADNGIYVIDETPAVGNLSLIGIFPAKGNPKELYSBAVNGETQ 495
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 427 AHKQAIRELJARDKNSAVVMWSIANEPASHEDGAREYFEPVLNLTROLDPTPTTANV 486
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 496 AHLOAIKELLARDQNHSSVVMWSIANEDPTPQ--VHGNISPLAETRLDPTPTTCVNV 554
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 487 GTATYQDLRIISLFDVSCINRYFGYISOTGLLEEAALKEKLEHGMOKFRPIYMTVEY 546
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 555 MFCDAHTDTISDLPTVCLNRYGYVOSGDIETKEKLEKELAMOKELHQPIITTEYG 614
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 547 ADTLAGHSILGLPMWSEFQVOMLDMVHRVFDRIISMAGEHVMNFPQTNLGIIRVDGN 606
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 615 VDTLAGHSMTDMWSEYQCAWLDMVHRVFDRIISMAGEHVMNFPQTNLGIIRVCGN 674
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 607 KKGVFTDRKRKAANAHSIRAWTSID 632
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 675 KKGIFTDRKRKPSAFLQKRWTKMN 700
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 15
US-10-389-640-23
Sequence 23, Application US/10389640
Publication No. US20040018591A1
GENERAL INFORMATION:
APPLICANT: Butt, Tauseef
APPLICANT: Weeks, Stephen
APPLICANT: Tran, Hiep
APPLICANT: Malakhova, Oksana
APPLICANT: Malakhov, Micheal
TITLE OF INVENTION: Methods and Compositions for Protein
FILE REFERENCE: 1955-P02972US2
CURRENT APPLICATION NUMBER: US/10/389,640
CURRENT FILING DATE: 2003-03-14
PRIOR APPLICATION NUMBER: 10/338,411
PRIOR FILING DATE: 2003-01-07
PRIOR APPLICATION NUMBER: 60/346,449
PRIOR FILING DATE: 2002-01-07
NUMBER OF SEQ ID NOS: 65
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 23

LENGTH: 711
TYPE: PRF
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic Sequence
US-10-389-640-23

Query Match 49.6%; Score 1665; DB 15; Length 711;
Best Local Similarity 52.4%; Pred. No. 2, 2e-132;
Matches 328; Conservative 98; Mismatches 178; Indels 22; Gaps 12;

QY 19 TPAARHFPNEMTO--HEOPL--IKVRPORTSSRELVLGLMKPAL--ASGLNDTAQ 70
DB 85 TPEDLDMDNDITIEAREQIGGMEFMLRPVETPTREIKKLDGLMAFSLDRENGCIGO--R 142
QY 71 PWTAPLPKGLCECPVPSYNDIFISREIHGVGVVYQREYIVPKGWSQERYLVRAESATH 130
DB 143 WMSALQESRAIAVPGSFNDQFADADIRNAGVWYQREVFIPKMGAGORIVLRFDAVTH 202
QY 131 HGRIVYNNRLVAEHVGGYTPFEADVTELVAPGKFRLLTGNNELTHETIPGKITTGNA 190
DB 203 YGKVMVNNQEWMEHQGYTPFEADVTPYIAGSVKITVCNNELMWOTIPGGMVIT--DE 261
QY 191 TGRRIQTYQHDFFYNAGLARSIMLVSPQOHIDITVVTDV--DGDGLINYEVEVANOT 248
DB 262 NGKKKQSYFPHDFENVAGIHRSVMLYTPPTWVDITVTVHVAQDCNHASVDMQV--VAN-- 318
QY 249 TGOIQISVIDEDGAIYAKASGAOGYTTISVKLMQGAATLYOLOVNIYSSGDVVDYTN 308
DB 319 -GDVSYELDADADQVAVATGQTSGLQVNPMLMQPBGGLYELCV--TAKSQTECDIYP 375
QY 309 LATGVTVRVAGSQFLINGKPFYFTGFGKHEDTAVRGKGDPAVYVHDFOLMKMGANSF 368
DB 376 LRVGIRSVAVKGOQFLINHKPFYFTGFGKHEDADLRKGFNDVLMVHDHALMDMIGANSY 435
QY 369 RFSHYPAEEMVDFADNRGIVVIDETPAVGLNIAL--MGVSESGAP--QFTTPDAINDKTOE 426
DB 436 RFSHYPAEEMVDFADNRGIVVIDETPAVGLNIAL--MGVSESGAP--QFTTPDAINDKTOE 495
QY 427 AHKQAIKELIARDKNTASVYVMSIANEPASHEDGAREFEPLTNLTROLDPTPTIFANY 486
DB 496 AHLQAIKELIARDKNTASVYVMSIANEPASHEDGAREFEPLTNLTROLDPTPTIFANY 554
QY 487 GATATVQLDRISDLFDVSCINRYFGWYSQSGDLSEAEALEKELHGOEKFRPIVTEYG 546
DB 555 MCDADHTDTISDLFDVCLNRYFGWYSQSGDLSEAEALEKELHGOEKFRPIVTEYG 614
QY 547 ADTLAHLHSILGLPMSEEFQVQMLDMYHRVFDRIESMAGEHWVNFADPQTNLGIIRVDGN 606
DB 615 VDTLAGLHSMYTDMSSEYQCAWLDMYHRVFDRIESMAGEHWVNFADPQTNLGIIRVDGN 674
QY 607 KKGVFTTRDRKPKAAHSLPARWTSID 632
DB 675 KKGIFTTRDRKPKSAFLQKRWTKMN 700

Search completed: March 18, 2005, 23:52:14
JOB time : 62 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2005, 23:45:19 ; Search time 43 Seconds
(without alignments)
1100.640 Million cell updates/sec

Title: US-10-757-093-4
Perfect score: 3354
Sequence: 1 MKFLGLSLSLAPSLGPN.....RKPAAASLRARWTSIDKN 634

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgnt2_6/prodata/1/iaa/5A_COMB.pep:*
2: /cgnt2_6/prodata/1/iaa/5B_COMB.pep:*
3: /cgnt2_6/prodata/1/iaa/6A_COMB.pep:*
4: /cgnt2_6/prodata/1/iaa/6B_COMB.pep:*
5: /cgnt2_6/prodata/1/iaa/6C_COMB.pep:*
6: /cgnt2_6/prodata/1/iaa/6D_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1721.5	51.3	832	3	US-08-630-820-7
2	1721.5	51.3	832	4	US-09-273-453-7
3	1714	51.1	1010	4	US-09-118-726-12
4	1711	51.0	603	3	US-09-149-727-6
5	1711	51.0	603	4	US-09-270-957-17
6	1711	51.0	603	4	US-09-270-957-23
7	1706	50.9	604	4	US-09-893-525-37
8	1706	50.9	659	4	US-09-893-525-40
9	1706	50.9	850	4	US-08-882-704-5
10	1664.5	49.6	602	2	US-08-882-704-5
11	1664.5	49.6	602	4	US-09-151-957-5
12	1664.5	49.6	602	6	5432081-2
13	1664.5	49.6	602	6	5432081-2
14	1664.5	49.6	1242	4	US-09-488-270A-2
15	1638.5	48.9	600	6	5268463-2
16	1638.5	48.9	600	6	5268463-2
17	1439	42.9	607	3	US-09-149-727-8
18	1439	42.9	607	4	US-09-270-957-18
19	1439	42.9	615	4	US-09-270-957-28
20	1439	42.9	618	3	US-09-149-727-4
21	1437	42.8	602	3	US-09-149-727-2
22	1437	42.8	602	4	US-09-270-957-2
23	1437	42.8	602	4	US-09-270-957-8
24	1437	42.8	602	4	US-09-270-957-15
25	1338	39.9	613	3	US-09-149-727-5
26	1338	39.9	613	4	US-09-270-957-16
27	1338	39.9	651	4	US-09-715-858-2

28	1323	39.4	648	4	US-09-715-858-4	Sequence 4, Appl1
29	1077.5	32.1	598	4	US-09-862-660-2	Sequence 2, Appl1
30	1027	30.6	500	4	US-09-949-016-11697	Sequence 11697, A
31	993.5	29.6	376	4	US-09-270-957-4	Sequence 4, Appl1
32	993.5	29.6	376	4	US-09-270-957-19	Sequence 19, Appl1
33	988	29.5	372	4	US-09-270-957-3	Sequence 3, Appl1
34	988	29.5	372	4	US-09-270-957-22	Sequence 22, Appl1
35	883	26.3	563	4	US-09-270-957-6	Sequence 6, Appl1
36	883	26.3	563	4	US-09-270-957-21	Sequence 21, Appl1
37	756.5	22.6	540	4	US-09-270-957-5	Sequence 5, Appl1
38	738	22.0	535	4	US-09-270-957-20	Sequence 20, Appl1
39	473.5	14.1	385	4	US-09-270-767-43634	Sequence 43634, A
40	413.5	12.3	1053	4	US-09-134-000C-5361	Sequence 5361, Ap
41	395.5	11.8	282	4	US-09-634-238-345	Sequence 345, App
42	390.5	11.6	1039	4	US-09-501-136-2	Sequence 2, Appl1
43	372.5	11.1	1010	4	US-09-654-448-2	Sequence 2, Appl1
44	372.5	11.1	1010	4	US-09-759-152A-2	Sequence 2, Appl1
45	372.5	11.1	1121	1	US-07-789-915A-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1
US-08-630-820-7
; Sequence 7, Application US/08630820
; Patent No. 6008023
; GENERAL INFORMATION:
; APPLICANT: OPPER, Martin
; APPLICANT: BOSSLET, Klaus
; APPLICANT: CZECH, Joerg
; TITLE OF INVENTION: CYTOPLASMIC EXPRESSION OF ANTIBODIES,
; TITLE OF INVENTION: ANTIBODY FRAGMENTS AND ANTIBODY FRAGMENT FUSION MOLECULES
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,820
; FILING DATE: 10-APR-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 19513676.4
; FILING DATE: 11-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: GRANADOS, Patricia D.
; REGISTRATION NUMBER: 33,683
; REFERENCE/DOCKET NUMBER: 18748/306
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; TELEX: 904136
; INFORMATION FOR SEQ. ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 832 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-630-820-7
Query Match 51.3%; Score 1721.5; DB 3; Length 832;
Best Local Similarity 52.9%; Pred. No. 2.4e-134;
Matches 339; Conservative 98; Mismatches 179; Indels 25; Gaps 11;

QY 8 SLTSLAAPSLSGTPA-----ARHPRNEMTOHEQPL-----IKVRPORTSSRELVLNDGLMK 58
 DB 190 SVTVVSSSLGTYTCVNNHKSNTKVDKRVBELSGSGSMWRPVETPREIKKLDGLWA 249
 QY 59 FAL---ASGLNTAQWTPAPLPGKLECPYPASYNDFISREIHDHVGWYQREVIYVPGK 115
 DB 250 FSLDRENCIGIDQ--RWMSALQESRAIAVPGSFNDQFADADIRNAGVNWYQREVEIYVPGK 307
 QY 116 WSGERLVPAESATTHGRITVNNRLVAEHVGGVTPPEADVTELVAEGEKRLTIGVNNEL 175
 DB 308 MAGQRLVLPFADVTHGKWNNOEVMHOGGTYTPPEADVTVYIAGKSVRLTVCVNNEL 367
 QY 176 THEITPCKITGNATGKRIQTYQHDVYVYAGIARSIMLYSVPOQHIODITVTVDV--DG 233
 DB 368 NMOTIPPGWVIT--DENGKKQSYFHFNFYVAGIHSVMLYTPNTVWDITVTVTHVAQDC 426
 QY 234 DNGLIYEVVANQTTGQIQISVIDEDGALVAKASAGCTVITPSVKLMQPGAAVLYQIQ 293
 DB 427 NHASVDMQV-VAN--GDVSVELRDADQGVATGQTSGLQVNVNPHLMQPGEGYLYELC 482
 QY 294 VNIIVSSGVDVDTYNLATGVRTYKVAQSQFLNGKPEYFTGFGKHEDTAVRGKCHDPAYM 353
 DB 483 V--TAKSQTECDIYPLRVGIRSVAVKGEQPLNHNKPFYFTGFGKHEDADLRGKGFNVLM 540
 QY 354 VHDFOIMKWIIGANSFRTSHYPAEEVMDPADRNGIYVIDETPAVGLNIAL--MGVSESGAP 412
 DB 541 VHDHALMDWIGANSYRTSHYPAEEMLDWADDEHGIYVIDETAANGNISLGIGFEAGNRP 600
 QY 413 -QTFPDANDKTOEAHKOAIRELIAARDKNHSAVVMWSIANEPASHEDGAREYFEPLTNL 471
 DB 601 KEIYSEAVANGETOQHLOAIKELIARDKNHPSVVMWSIANEBDTPQGRAREFAFALAA 660
 QY 472 TROLDPRTPTFANVTATYQDLRISDLPVSCINRYFGMYSGTGLBEAEALKEKELHG 531
 DB 661 TKLDPRTPTICVNVAFCDAHDTITSDLPVLCINRYGYVQSGLLEKLEKLELLA 720
 QY 532 WOEKFRPIVMTTEYAGDTLAGHSILGLPWESEFOVQMLDMYHVRVEDRIESNAGEHVNF 591
 DB 721 WOEKQHQPITITREYGVDTLAGHSMTYDMSSEYOCAMLDMYHVRVEDRVSAVVGEQVWVF 780
 QY 592 ADFQTNLGIIRDGNKKGVFTDRDKPKAAHSLRAWTSID 632
 DB 781 ADFATSOGILRVGNKKGIIFTRDRKPKSAFLLQKRWTKMN 821
 RESULT 2
 US-09-273-453-7
 / Sequence 7, Application US/09273453
 / Patent No. 6602688
 / GENERAL INFORMATION:
 / APPLICANT: Oppner, Martin
 / BOSSLER, Klaus
 / CZECH, Joerg
 / TITLE OF INVENTION: CYTOPLASMIC EXPRESSION OF ANTIBODIES,
 / ANTIBODY FRAGMENTS AND ANTIBODY FRAGMENT FUSION MOLECULES
 / IN E. COLI
 / NUMBER OF SEQUENCES: 7
 / CORRESPONDENCE ADDRESS:
 / ADDRESSEE: Foley & Lardner
 / STREET: 3000 K Street, N.W., Suite 500
 / CITY: Washington
 / STATE: D.C.
 / COUNTRY: USA
 / ZIP: 20007-5109
 / COMPUTER READABLE FORM:
 / MEDIUM TYPE: Floppy disk
 / COMPUTER: IBM PC compatible
 / OPERATING SYSTEM: PC-DOS/MS-DOS
 / SOFTWARE: Patentin Release #1.0, Version #1.30
 / CURRENT APPLICATION DATA: US/09/273,453
 / FILING DATE: 22-Mar-1999
 / PRIOR APPLICATION DATA:

/ APPLICATION NUMBER: 08/630,820
 / FILING DATE: <Unknown>
 / ATTORNEY/AGENT INFORMATION:
 / NAME: GRAMADOS, Patricia D.
 / REGISTRATION NUMBER: 33,683
 / REFERENCE/DOCKET NUMBER: 18748/306
 / TELECOMMUNICATION INFORMATION:
 / TELEPHONE: (202)672-5300
 / TELEFAX: (202)672-5399
 / TELE: 904136
 / INFORMATION FOR SEQ ID NO: 7:
 / SEQUENCE CHARACTERISTICS:
 / LENGTH: 832 amino acids
 / TYPE: amino acid
 / TOPOLOGY: linear
 / MOLECULE TYPE: protein
 / SEQUENCE DESCRIPTION: SEQ ID NO: 7:
 / US-09-273-453-7
 Query Match 51.3%; Score 1721.5; DB 4; Length 832;
 Best Local Similarity 52.9%; Pred. No. 2.4e-134;
 Matches 339; Conservative 98; Mismatches 179; Indels 25; Gaps 11;
 QY 8 SLTSLAAPSLSGTPA-----ARHPRNEMTOHEQPL-----IKVRPORTSSRELVLNDGLMK 58
 DB 190 SVTVVSSSLGTYTCVNNHKSNTKVDKRVBELSGSGSMWRPVETPREIKKLDGLWA 249
 QY 59 FAL---ASGLNTAQWTPAPLPGKLECPYPASYNDFISREIHDHVGWYQREVIYVPGK 115
 DB 250 FSLDRENCIGIDQ--RWMSALQESRAIAVPGSFNDQFADADIRNAGVNWYQREVEIYVPGK 307
 QY 116 WSGERLVPAESATTHGRITVNNRLVAEHVGGVTPPEADVTELVAEGEKRLTIGVNNEL 175
 DB 308 MAGQRLVLPFADVTHGKWNNOEVMHOGGTYTPPEADVTVYIAGKSVRLTVCVNNEL 367
 QY 176 THEITPCKITGNATGKRIQTYQHDVYVYAGIARSIMLYSVPOQHIODITVTVDV--DG 233
 DB 368 NMOTIPPGWVIT--DENGKKQSYFHFNFYVAGIHSVMLYTPNTVWDITVTVTHVAQDC 426
 QY 234 DNGLIYEVVANQTTGQIQISVIDEDGALVAKASAGCTVITPSVKLMQPGAAVLYQIQ 293
 DB 427 NHASVDMQV-VAN--GDVSVELRDADQGVATGQTSGLQVNVNPHLMQPGEGYLYELC 482
 QY 294 VNIIVSSGVDVDTYNLATGVRTYKVAQSQFLNGKPEYFTGFGKHEDTAVRGKCHDPAYM 353
 DB 483 V--TAKSQTECDIYPLRVGIRSVAVKGEQPLNHNKPFYFTGFGKHEDADLRGKGFNVLM 540
 QY 354 VHDFOIMKWIIGANSFRTSHYPAEEVMDPADRNGIYVIDETPAVGLNIAL--MGVSESGAP 412
 DB 541 VHDHALMDWIGANSYRTSHYPAEEMLDWADDEHGIYVIDETAANGNISLGIGFEAGNRP 600
 QY 413 -QTFPDANDKTOEAHKOAIRELIAARDKNHSAVVMWSIANEPASHEDGAREYFEPLTNL 471
 DB 601 KEIYSEAVANGETOQHLOAIKELIARDKNHPSVVMWSIANEBDTPQGRAREFAFALAA 660
 QY 472 TROLDPRTPTFANVTATYQDLRISDLPVSCINRYFGMYSGTGLBEAEALKEKELHG 531
 DB 661 TKLDPRTPTICVNVAFCDAHDTITSDLPVLCINRYGYVQSGLLEKLEKLELLA 720
 QY 532 WOEKFRPIVMTTEYAGDTLAGHSILGLPWESEFOVQMLDMYHVRVEDRIESNAGEHVNF 591
 DB 721 WOEKQHQPITITREYGVDTLAGHSMTYDMSSEYOCAMLDMYHVRVEDRVSAVVGEQVWVF 780
 QY 592 ADFQTNLGIIRDGNKKGVFTDRDKPKAAHSLRAWTSID 632
 DB 781 ADFATSOGILRVGNKKGIIFTRDRKPKSAFLLQKRWTKMN 821
 RESULT 3
 US-09-118-276-12
 / Sequence 12, Application US/09118276
 / Patent No. 6693185
 / GENERAL INFORMATION:

APPLICANT: BABYCHUK, ELENA;
 APPLICANT: KUSHNIR, SERGEI;
 APPLICANT: DE BLOCK, MARC;
 APPLICANT: INZE, DIRK
 TITLE OF INVENTION: METHODS AND MEANS TO MODULATE PROGRAMMED
 DEATH IN EUKARYOTIC CELLS
 NUMBER OF SEQUENCES: 21
 CORRESPONDENCE ADDRESS:
 ADDRESS: SIXEY, FRIEDMAN, LEEDOM, & FERGUSON
 STREET: 8180 GREENSBORO DRIVE, SUITE 800
 CITY: MCLEAN,
 STATE: VIRGINIA
 COUNTRY: USA
 ZIP: 22102
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3-1/2" DISKETTE
 COMPUTER: IBM-COMPATIBLE
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: ASCII
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/118,276
 FILING DATE: 17-JUL-1998
 PRIOR APPLICATION DATA: NONE
 ATTORNEY/AGENT INFORMATION:
 NAME: SCHULMAN, ROBERT M.; SALEN, KENNETH H.
 REGISTRATION NUMBER: 31,196; 43,077
 REFERENCE/DOCKET NUMBER: 6201-0003
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 790-9110
 TELEFAX: (703) 883-0370
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1010 RESIDUES
 TYPE: AMINO ACID
 STRANDEDNESS: SINGLE
 TOPOLOGY: LINEAR
 US-09-118-276-12

Query Match 51.1%; Score 1714; DB 4; Length 1010;
 Best Local Similarity 55.2%; Pred. No. 1,4e-133;
 Matches 331; Conservative 92; Mismatches 161; Indels 16; Gaps 9;
 40 VRPQRTSRELVLNDGLMKPAL--ASGLNDTAQPTAPLPGKLECPVASYNDIFISRE 96
 409 VRPVEPTREIKKLDGLMAFSLDRENCIDQ--RWMESALOESRAIAVPSFNDQFADAD 466
 97 IHDHVGMYVYQREVIYVPGKWSQERLYVRAESATHGRIYVNNRLVAEHVGGYTPFEADVT 156
 467 IRNYAGNVWYQREVFIPKMGAGORIVLRPDVATHYGVWVNNQEVHEHGGYTPFEADVT 526
 157 ELVAPGEKRLITGVNNELTHEITIPPKITGTGATGKRIOTYQHDFFNYAGLARISWLYS 216
 527 PYVIAKGSVRIITVCNNELMWQITIPGMVIT--DENGKKQSYFHPDFNYAGIHSVWLYT 585
 217 VPQOHODITVVDV--DGNGLINYEVEYANQTOGOIOSVIDEGAIYAKXSGAAGTV 274
 586 TPTWVDDITVTVHVAODCNHASVDWQV--VAN--GDVSYELDADQOVVATGCGTSGL 641
 275 TIPSVKLMQGAAYLYQLQVNIIVSGSDVDVYTNLATGVRTVYVAGSQPLINKPFFYTG 334
 642 QVNNPHLMQGBEGYLVELCV--TAKSQTECDIYPLRGIRSVAVKSGQPLINKPFFYTG 699
 335 FCGHEDTAVRGKGDPAVYVNDPOLMKWIGANSFRTSHYVYAEVWDFADRNGIVVDET 394
 700 FCGHEDADLRGKGFVDVLMVHDALMDWIGANSYRTSHYVYAEVWDFADRNGIVVDET 759
 395 PAYGLNIAL--MGVSESGAP--QTFTPDAINDKTOEAKHQAIRELIARDKXNASVVMMSIAN 452
 760 AAVGFNLISLIGEGAGNKPKELYSEEVNGETOQAHQALKEILARDKXNPVVMMSIAN 819
 453 EPASHEDGAREYEPFLNLTROLDPTRPITFANVGATYQLDRIISDLFDVSCINRYFGWY 512
 820 EPDTRRGAREYFAPLAELATRKLDPTRPITCVNVVPCDANTDTISDLFDVLCINRYFGWY 879

RESULT 4
 US-09-149-727-6
 ; Sequence 6, Application US/09149727
 ; Patent No. 6391547
 ; GENERAL INFORMATION:
 ; APPLICANT: Jefferson, Richard A.
 ; APPLICANT: Kilian, Andrej
 ; APPLICANT: Keese, Paul Konrad
 ; TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND
 ; FILE REFERENCE: 190106 405
 ; CURRENT APPLICATION NUMBER: US/09/149,727
 ; CURRENT FILING DATE: 1998-09-08
 ; EARLIER APPLICATION NUMBER: US 60/058,263
 ; EARLIER FILING DATE: 1997-09-09
 ; NUMBER OF SEQ ID NOS: 71
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 6
 ; LENGTH: 603
 ; TYPE: PRT
 ; ORGANISM: Escherichia coli
 US-09-149-727-6

Query Match 51.0%; Score 1711; DB 3; Length 603;
 Best Local Similarity 55.0%; Pred. No. 1.1e-133;
 Matches 330; Conservative 93; Mismatches 161; Indels 16; Gaps 9;
 40 VRPQRTSRELVLNDGLMKPAL--ASGLNDTAQPTAPLPGKLECPVASYNDIFISRE 96
 2 LRPEVPTREIKKLDGLMAFSLDRENCIDQ--RWMESALOESRAIAVPSFNDQFADAD 59
 97 IHDHVGMYVYQREVIYVPGKWSQERLYVRAESATHGRIYVNNRLVAEHVGGYTPFEADVT 156
 60 IRNYAGNVWYQREVFIPKMGAGORIVLRPDVATHYGVWVNNQEVHEHGGYTPFEADVT 119
 157 ELVAPGEKRLITGVNNELTHEITIPPKITGTGATGKRIOTYQHDFFNYAGLARISWLYS 216
 120 PYVIAKGSVRIITVCNNELMWQITIPGMVIT--DENGKKQSYFHPDFNYAGIHSVWLYT 178
 217 VPQOHODITVVDV--DGNGLINYEVEYANQTOGOIOSVIDEGAIYAKXSGAAGTV 274
 179 TPTWVDDITVTVHVAODCNHASVDWQV--VAN--GDVSYELDADQOVVATGCGTSGL 234
 275 TIPSVKLMQGAAYLYQLQVNIIVSGSDVDVYTNLATGVRTVYVAGSQPLINKPFFYTG 334
 235 QVNNPHLMQGBEGYLVELCV--TAKSQTECDIYPLRGIRSVAVKSGQPLINKPFFYTG 292
 335 FCGHEDTAVRGKGDPAVYVNDPOLMKWIGANSFRTSHYVYAEVWDFADRNGIVVDET 394
 293 FCGHEDADLRGKGFVDVLMVHDALMDWIGANSYRTSHYVYAEVWDFADRNGIVVDET 352
 395 PAYGLNIAL--MGVSESGAP--QTFTPDAINDKTOEAKHQAIRELIARDKXNASVVMMSIAN 452
 353 AAVGFNLISLIGEGAGNKPKELYSEEVNGETOQAHQALKEILARDKXNPVVMMSIAN 412
 453 EPASHEDGAREYEPFLNLTROLDPTRPITFANVGATYQLDRIISDLFDVSCINRYFGWY 512
 413 EPDTRRGAREYFAPLAELATRKLDPTRPITCVNVVPCDANTDTISDLFDVLCINRYFGWY 472
 513 SGTGDLBEAALAEKELHGWQEFKRPVYVTEYAGDTLAGLHSLILGPMSEEVQVOMLDM 572
 473 VQSGDLETAKEVLEKELHGWQEFKRPVYVTEYAGDTLAGLHSLILGPMSEEVQVOMLDM 532

OY	YHRVDRIRIESAGEHHVMNPADEQTNIGIIRDGNKKGVETPRRKPAAASHLRAPWTSID	6332
573	:::~::~:	:
Ddb	YHVRDVSAAVGEQVMNFAPATSCILRGANKKGIPTRRKKPXSAAFLLQIKRWGMN	5922
533	:::~::~:	:

RESULT 5

```

US-09-270-957-17
; Sequence 17, Application US/09270957
; Patent No. 6641996
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL - GLUCURONIDASE GENES, GENE
; TITLE OF INVENTION: PRODUCTS, AND USES THEREOF
; FILE REFERENCE: 190106, 405C1
; CURRENT APPLICATION NUMBER: US/09/270,957
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 603
; TYPE: prt
; ORGANISM: Escherichia coli
; US-09-270-957-17

```

Query Match	51.0%	Score 1711	DB 4	Length 603	
Best Local Similarity	55.0%	Pred. No. 1.1e-133			
Matches 330	Conservative 93	Mismatches 161	Indels 16	Gaps 9	

[illegible]

RESULT 6

US-09-270-957-23
; Sequence 23, Application US/09270957
; Patent No. 6641996

```

; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL -GLUCURONIDASE GENES, GENE
; TITLE OF INVENTION: PRODUCTS AND USES THEREOF
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/09/270,957
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
;
; LENGTH: 603
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-270-957-23

```

Query Match	51.0%;	Score 1711;	DB 43;	Length 603;
Best Local Similarity	55.0%;	Pred. No. 1,1e-133;		
Matches 330;	Conservative 93;	Mismatches 161;	Indels 16;	Gaps 97;

Qy	40	VRPQRISSEELVNLIDOLMFEAL--ASGINDPAQWTPLEKJCECPAPASVINDIFPSRE	96
Db	2	LREVEFTPEIKKLDOALFSLDRENCIDQ--RMESALDESAIIVPSSFNDFQADAD	59
Qy	97	IHDVGVMYVYOREBVIIVPKMSQERYLVAESAITHGRILYVNRULVAEIVGGYTFEADVT	156
Db	60	IRYAVGNVWYQGEVFLPKMGAGQRIYLRFDAVTHGKVMVNNQGEVMEHGGSYTFEADVT	119
Qy	157	ELVAPEEKRLITGVNNELTETTPPKKITGNATGKRIQYQYDHPNYAGLARIWLYS	216
Db	120	PVYIAKSKSVRIITVCVNNELMQITIPPQVWIT--DENGKKKQSYFHDFFNYAGIHRSVWL,YT	178
Qy	217	VPPQHIQDITLVTDV--DGDNGLIYEVYEVANQTTGQIQISVIDEDGAIYAKKSGAQGV	274
Db	179	TPWTWDDITLVTHVAQDNCNHSVMOY-VAN--GDVSELRADQOVATGGTISGTL	234
Qy	275	TIPSVKLWOPGAAYLYLOLVNVGVSSGGVVDTYMLATGVRIVKAGSQFLINKRPFYFTG	334
Db	235	QVNVPHLMQPGSGYLVELCV--TASQTECOIPLRNGIRBVAVAKGQFLINHPRFFFTG	292
Qy	335	FGHEEDTAVRGKGHDPAYWHDFOLMKRTIGANSFRTSHYPAAEVMDPADBNIGIVDET	394
Db	293	FGHEEDADIRGKGFQDVILVWHDHMLMDWIGANSYSTSHYPAAEMLMDADEHGIVDET	352
Qy	395	PARGNLIA--MCVSESGAP--OTPEFADINDTQEAHKAITELIARDKNHSVVMMSIAN	452
Db	353	AAGFNPLSGIGFEGKNGKPELSEAYNGEIOQHLDALIKELIARDKNHPSVVMMSIAN	412
Qy	453	EPASHEDGAREYFEPILNLTROLDTREITFEANVTATYQLDRIISDLFDVSCINRYGWY	512
Db	413	EPDTRQGARREYFAPLAEXTRLDPTTRITICVNMFCDAHNDTISDLFDVLICNRYGWY	472
Qy	513	SQTGDLSEBAALEKELHGMQEKFRPLVMTREYADTLAHSITLGLPMSEEPQVQMLDM	572
Db	473	VQSGDLETAEKLEKELLAMOEKHLQPIITIEYGVDTLAGHSYTMDSMEEYQCAWLDM	532
Qy	573	YHHPVPRISMAGEIVMNPADPQNLGIIRVNDKKVGVFRRDRKPKAAHSLAARPTSID	632
Db	533	YHHPVPRVAVGBOVMNPADPAISOGILIRGVGNKKGIIFTRDRKPKSAADLLKRPMTGN	592

RESULT 7

```

US-09-893-525-37
; Sequence 37, Application US/09893525
; Patent No. 6753167
;
; GENERAL INFORMATION:
; APPLICANT: Koloney, Maurice M.
; APPLICANT: Van Roijen, G1s
; TITLE OF INVENTION: Preparation of Heterologous Proteins on Oil Bodies
; FILE REFERENCE: 9369-172
; CURRENT APPLICATION NUMBER: US/09/893,525
; CURRENT FILING DATE: 2001-06-28
; PRIORITY APPLICATION NUMBER: US 09/210,843
; PRIORITY FILING DATE: 1998-12-15

```



```

? PRIOR APPLICATION NUMBER: US 08/846,021
? PRIOR FILING DATE: 1997-04-25
? PRIOR APPLICATION NUMBER: US 08/366,783
? PRIOR FILING DATE: 1994-12-30
? PRIOR APPLICATION NUMBER: US 08/142,418
? PRIOR FILING DATE: 1993-11-16
? PRIOR APPLICATION NUMBER: US 07/659,835
? PRIOR FILING DATE: 1991-02-22
? NUMBER OF SEQ ID NOS: 42
? SOFTWARE: PatentIn version 3.1
? SEQ ID NO 37
? LENGTH: 604
? TYPE: PRT
? ORGANISM: Artificial Sequence
? FEATURES:
? OTHER INFORMATION: Phas-GUS-phas
US-09-893-525-37

```

Query Match	50.9%	Score 1706	DB 4	length 604
Best Local Similarity	54.8%	Pred. No. 2.9e-133		
Matches 329	Conservative 94	Mismatches 161	Indels 16	Gaps 9

Qy	40	VRPQTSRSLRELVNLDGIMKFAI--ASGLNDTPAQWPAIPLPKGLECPVPASYNDIFISRE	96
Db	3	LRPEVETPRTREIKKLDGLWAFSLDRENCISIDQ--RMMESALQESRAIAVPGSSFNDFPADAD	60
Qy	97	IHDVGVWVYQREYIVPKGMSQERVLVAESAETHGRIVYNNRLVAEYVGGYTPREADVT	156
Db	61	IRNAGVNWYQREYFIPKGMAGQRIVLNFPDAETHGKGVWVNWQEWMEHOGGTTPEADVT	120
Qy	157	ELVAPEGKFLITIGVNNELTHEITPIPGKITTGNAIYQIOTYQHPYVYAGLARSIMYLS	216
Db	121	PVYIAGKSVRITVCNNELNMQIPIPGWVIT--DENGKKKKQSYFHDFPNVYAGHSVIMLYT	179
Qy	217	VPGQIHODITVYTDV--DGDNGILNYEVEVANQTTGGQIQIYSYIDBDGAIKAKASQAQGV	274
Db	180	TPNFWVDDITVVTVAHQAODCNHASWQWY-VAN---GDVSVELRDADQCVAAVAGQSTGTL	235
Qy	275	TIPSVKLMQPGAAVYLQLOVNNIVGSSGSDVPDTPNATATGRTATKVGASQOFLINGKPFYFG	334
Db	236	QVNVPHLMQPEGGYIYELCV--TAKSQTECDIYPLRVGIRSAVAGQOFLINHKKPFYFTG	293
Qy	335	FGKHEDTAVRKGKNDPAYVHDFOLMKIIGANSFRTSHYPVAEEVWDFADRNGLIVIDET	394
Db	294	FGRHEDADLRKGKGFNVVHVDHALMDIIGANSYTSHPVAEEMLDWADDEHGIYVIDET	353
Qy	395	PAYGLNIAL-MGVSESGAP-QTFPDDAINDKTQEHKQAIIRLIRARDKSHVWVMSIIN	452
Db	354	AAVAFSSSLGIGFEAGNPKRELVEEAVNGEFOQHLOAIKELIARDKNHPVWMSIIN	413
Qy	453	EPASHBQAREYFELTNLTROLDPTRTPTFPAVNGTATVQDLRISDLFPVSGCINRNYFGY	512
Db	414	EPDTRPGQAREYFAPLIAETRTKLDPTRTIITCVNWNFCDAHNTDTSIDLFPVCLNRIYGY	473
Qy	513	SQTGDLEAEALAEKELHGWQEKPHRPIVMTGYGADTLAGHSISLGLPWSBEFYQYMDM	572
Db	474	VQSGDLETAEKYLEKELLAMQEKLHQPRIITEYGVDTLAGHSMTYDMMSEYQCAMLDM	533
Qy	573	YHRFPRDRIESAGHVNWFPADPQTNLGIIRDNGKKGVPTRPRKKAALHSIRAWTSID	632
Db	534	YHRFPRDRSAVAGQVWNPADPATSQGLIRVGNNKKGIFTRBRKKSAAFLQDKWGTGN	593

```

RESULT 8
US-09-893-525-40
; Sequence 40, Application US/09893525
; Patent No. 6753167
;
; GENERAL INFORMATION:
; APPLICANT: Moloney, Maurice M.
; APPLICANT: Van Rooijen, Gjs
; TITLE OF INVENTION: Preparation of Heterologous Proteins on Oil Bodies
; FILE REFERENCE: 9365-112
; CURRENT APPLICATION NUMBER: US/09/893,525

```

```

: CURRENT FILING DATE: 2001-06-29
: PRIOR APPLICATION NUMBER: US 09/220,843
: PRIOR FILING DATE: 1998-12-15
: PRIOR APPLICATION NUMBER: US 08/846,021
: PRIOR FILING DATE: 1997-04-25
: PRIOR APPLICATION NUMBER: US 08/366,783
: PRIOR FILING DATE: 1994-12-30
: PRIOR APPLICATION NUMBER: US 08/142,418
: PRIOR FILING DATE: 1993-11-16
: PRIOR APPLICATION NUMBER: US 07/659,835
: PRIOR FILING DATE: 1991-02-22
: NUMBER OF SEQ ID NOS: 42
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 40
: LENGTH: 659
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: phae-oleo GUS-phae
US-09-893-525-40

```

Query Match	50.9%	Score 1706;	DB 4;	length 659;
Best Local Similarity	54.8%;	Pred. No. 3	3e-133;	
Matches 329;	Conservative	94;	Mismatches 161;	Indels 16;
				Gaps 9;

QY	40	VROQRTSSREL.VNLDGIMKFL--ASGLNDTQAPMTAPLPKIECEVPVASYNDITFSRE	96
Db	58	LRREVEFTREIKKLDGLMAFSLDRENGCIGD--RWMSEALOESPAIVPFGSFNOQFADAD	115
QY	97	IHDHVGVVYOREEVIYVPKGMSOERYIVARASATHHGIIYNNRLVAHEVGATPEFADVT	156
Db	116	IRNYAGVWYOREEFTPIPKGWAQCIYLRFDVATHYGVWANNQDVMHEHOGGYTFPEADVT	175
QY	157	ELVAPGKFFRLITGVNNELTHEETIPKGIITTGATGKRIQTYOHDFYNYAGLARSIMLS	216
Db	176	PYIYAGSVAITVCANNELMQGITPEGMVIT--DENGKKQSYFADFNYAGIHRSVMLYT	234
QY	217	VPOQHODITIVTVDV--DGDGLNIEVEYEAANOQTGQIQISVIDEDGALVAKASGAQCTV	274
Db	235	TPMTWDDITIVVTHVADQCNHASDVMGV--VAN--GVVSVELRDAOOVAATGGTSGTL	290
QY	275	TIPSVKLMORGAAVLYXOLOYNIWSSSDVDVDTYLAATGVTPVKAAGSOFGLNGKPFYTG	334
Db	291	QVNNPHLMQEGEYIELCV--TAKSGTECDIYPLRGISNSAVAKGQOFLINHKEPFYTG	348
QY	335	FGKHEDTAVAGKGDPAVMVHDFOLMKIGANSFRTSHYPAAEVMDFADNGIIVIDET	394
Db	349	FGRHEDADLKGKGFVDVYVHDLMDWIMIGANSYRTSHYPAAEEMLDWADDEHGIIVIDET	408
QY	395	PAYGNAL--MGVSESGAP--QTFRPDALNDKTOEAAHQALRELLARQDNHASYVMWSIAN	452
Db	409	AAVGFSLDIOIGEPAGKPKELYSEEVNETOQAHLOAKELIARQDNHPSVVMWSIAN	468
QY	453	EPASHEDGAEYEPEPLTNLTROLDPTPARIFFANYGTAITYOLDRIISDLFEDVSCINRYEGMY	512
Db	469	EPDTRPGGAEYEPALAEATRKLDPTPRTICVAMPCDATHDTISDLFEDVLCNRYEGMY	528
QY	513	SQTGDLEAEALAEKELHGWQEKFRHDYVMTGYADTLAAGHSITLGLPMSEEFQVMXMDM	572
Db	529	VQSGDLETAKEVLEKELLAQOEKIHQPIITTEYGVDTLAGHSYTYMMSSEYOCAMXMDM	588
QY	573	YHVFEDIEEMAGEHNVPADPQNIIGIITVDNGKGVFPRDRKPKXAASHLRBMTSID	632
Db	589	YHRFEDVSVAVVSGQVNNPDAFATISQGLIRGGKKKQIFTRDRKPKXSAFLQCKRMGMN	648

RESULT 9
US-09-893-525-42
; Sequence 42, Application US/09893525
; Patent No. 6753167
; GENERAL INFORMATION:
; APPLICANT: Moloney, Maurice M.
; APPLICANT: Van Rooijen, Gijb

TITLE OF INVENTION: Preparation of Heterologous Proteins on Oil Bodies
FILE REFERENCE: 9369-1172
CURRENT APPLICATION NUMBER: US/09/893,525
CURRENT FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 09/210,843
PRIOR FILING DATE: 1998-12-15
PRIOR APPLICATION NUMBER: US 08/846,021
PRIOR FILING DATE: 1997-04-25
PRIOR APPLICATION NUMBER: US 08/366,783
PRIOR FILING DATE: 1994-12-30
PRIOR APPLICATION NUMBER: US 08/142,418
PRIOR FILING DATE: 1993-11-16
PRIOR APPLICATION NUMBER: US 07/659,835
PRIOR FILING DATE: 1991-02-22
NUMBER OF SEQ ID NOS: 42
SOFTWARE: PatentIn version 3.1
SEQ ID NO 42
LENGTH: 850
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: phas-caleo-GUS-phas
US-09-893-525-42

Query Match 50.9%; Score 1706; DB 4; Length 850;
Best Local Similarity 54.8%; Pred. No. 4,9e-133;
Matches 329; Conservative 94; Mismatches 161; Indels 16; Gaps 9;
40 VRPQRSSRELVLNDGLMKFAL--ASGLNDIAQPMTPALPKGLCPYPASVNDIFISRE 96
249 LRPEVETREIKKLDGLMAFSLDRENGCIGD--RWMSALQESRALAVPESFNDGFADND 306
97 IHDHVGWVYQREVEIVPKGMSQERYLVRAESATHHGRIVYNNRLVAEHVGCTPFEADVT 156
307 IRVYAGWVYQREVEIVPKGMAQGRIVLRPDVTHVGKVVWVNOGVNEHOGGYTPFEADVT 366
157 ELVAPGEKFLITGVNNELTHEITPPGKITTGNAATGRIQTOYHDFPNYAGLARSTWLYS 216
367 PYVIAGKSVRITVCVNNELMWQITPPGMVIT--DENGKKKQSYFHDFPNYAGIHRSVMLYT 425
217 VPQOHODITVTVTV--DGDNGLINYEVEVANQTTGOIQISVIDEDGAIYAKASGAQGTV 274
426 TPTNWVDDITVTVHVAQDCHNASVDQV--VAN--GDVSELRLDADQGVVATGQGTSGTL 481
275 TIPSVKLMQGAAYLYQLOVNIYSSGDDVDTYNIATGVRYKVASQGLINGKPEYFTG 334
482 QVNNPHLMQGEGLYELCV--TAKSQTECDIYPLRVGIRSVAVKGGQFLINHKPEYFTG 539
335 FGKHEDTAVRGKGDPAVYVHDFOIMKVIIGANSFRTSHYPVAEVMDPADRNGIIVIDET 394
540 FGRHEDADLRGKGFDPVNLMDHMLDMIGANSYRTSHYPVAEVMDPADRNGIIVIDET 599
395 PAVGINIAL-MGVSESGAP-QTFTPDALINDKTOEAKHQAIRELIARDKXNASVVMMSIAN 452
600 AAVGFNSLIGIFGKAGKPEKELYSEAVNGETOQALQAIKELIARDKXNASVVMMSIAN 659
453 EPASHEDGAREYFEPITNLTRQDPTFRITFANVTATYQLDRIISDLFVSCINRKYGMV 512
660 EBDTRPQARREYFALAEATRKLDPTFRITFCVNVVFCDAHTDTISDLFDVLCLNRYGMV 719
513 SOTGLLEAEALAELELHGMQEKFRPIVMTVEYGAADTLAHLISLIGLPWSEEFQVOMLDM 572
720 VQSGULETAEKLEKELAMQEKHQPITITIEYGVDTLAGHSMVTIDMWSESEYQOAMLDM 779
573 YHRVEDRIEESMAGEHVNNFADFQTNLGIIRVDGNKKGVFTTRDRKXKAAASHLRARWISID 632
780 YHRVEDRISAVVGEQVNNFADFATQGLIRVGNKKGIIFTRDRKXKAAASHLRARWISID 839

RESULT 10
US-08-882-704A-5
Sequence 5, Application US/08882704A
Patent No. 5879906

GENERAL INFORMATION:
APPLICANT: Jefferson, Richard A.
APPLICANT: Wilson, Katherine J.
APPLICANT: Leader, Michael
TITLE OF INVENTION: GLUCURONIDE REPRESSORS AND USES THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESS: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/882,704A
FILING DATE: 25-JUN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 5879906tendburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 190106.404
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 602 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-882-704A-5

Query Match 49.6%; Score 1664.5; DB 2; Length 602;
Best Local Similarity 54.0%; Pred. No. 8,2e-130;
Matches 324; Conservative 93; Mismatches 166; Indels 17; Gaps 10;
40 VRPQRSSRELVLNDGLMKFAL--ASGLNDIAQPMTPALPKGLCPYPASVNDIFISRE 96
2 LRPEVETREIKKLDGLMAFSLDRENGCIGD--RWMSALQESRALAVPESFNDGFADND 59
97 IHDHVGWVYQREVEIVPKGMSQERYLVRAESATHHGRIVYNNRLVAEHVGCTPFEADVT 156
307 IRVYAGWVYQREVEIVPKGMAQGRIVLRPDVTHVGKVVWVNOGVNEHOGGYTPFEADVT 366
157 ELVAPGEKFLITGVNNELTHEITPPGKITTGNAATGRIQTOYHDFPNYAGLARSTWLYS 216
367 PYVIAGKSVRITVCVNNELMWQITPPGMVIT--DENGKKKQSYFHDFPNYAGIHRSVMLYT 425
217 VPQOHODITVTVTV--DGDNGLINYEVEVANQTTGOIQISVIDEDGAIYAKASGAQGTV 274
426 TPTNWVDDITVTVHVAQDCHNASVDQV--VAN--GDVSELRLDADQGVVATGQGTSGTL 481
275 TIPSVKLMQGAAYLYQLOVNIYSSGDDVDTYNIATGVRYKVASQGLINGKPEYFTG 334
482 QVNNPHLMQGEGLYELCV--TAKSQTECDIYPLRVGIRSVAVKGGQFLINHKPEYFTG 539
335 FGKHEDTAVRGKGDPAVYVHDFOIMKVIIGANSFRTSHYPVAEVMDPADRNGIIVIDET 394
540 FGRHEDADLRGKGFDPVNLMDHMLDMIGANSYRTSHYPVAEVMDPADRNGIIVIDET 599
395 PAVGINIAL-MGVSESGAP-QTFTPDALINDKTOEAKHQAIRELIARDKXNASVVMMSIAN 452
600 AAVGFNSLIGIFGKAGKPEKELYSEAVNGETOQALQAIKELIARDKXNASVVMMSIAN 659
453 EPASHEDGAREYFEPITNLTRQDPTFRITFANVTATYQLDRIISDLFVSCINRKYGMV 512
660 EBDTRPQARREYFALAEATRKLDPTFRITFCVNVVFCDAHTDTISDLFDVLCLNRYGMV 719
513 SOTGLLEAEALAELELHGMQEKFRPIVMTVEYGAADTLAHLISLIGLPWSEEFQVOMLDM 572
720 VQSGULETAEKLEKELAMQEKHQPITITIEYGVDTLAGHSMVTIDMWSESEYQOAMLDM 779
573 YHRVEDRIEESMAGEHVNNFADFQTNLGIIRVDGNKKGVFTTRDRKXKAAASHLRARWISID 632
780 YHRVEDRISAVVGEQVNNFADFATQGLIRVGNKKGIIFTRDRKXKAAASHLRARWISID 839

QY 513 SQTGDEAEBALEKEHMOEKFRPIVWTEYGADTLAHLISLGLPMSSEFOVQMLDM 572
 DB 472 VQSGDELEAKVLEKELAMQEKHQPIITTEYGVDTLGLHSMYTDMSSEYQCAMLDM 531
 QY 573 YHRVPRISMAGEHYWNPADPOTNIGIIRVDGNKKGVFTDRKPKAAHSLPARFTSID 632
 DB 532 YHRVPRISAVAGEQVWNPADPOTNIGIIRVDGNKKGVFTDRKPKAAHSLPARFTSID 591

RESULT 11

US-09-151-957-5
 ; Sequence 5, Application US/09151957
 ; Patent No. 6429292

GENERAL INFORMATION:
 APPLICANT: Jefferson, Richard A.
 Wilson, Katherine J.

Leader: Michael

TITLE OF INVENTION: GLUCURONIDE REPRESSORS AND USES THEREOF

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
 FILING DATE: 11-Sep-1998
 APPLICATION NUMBER: US/09/151,957
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/882,704
 FILING DATE: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: No. 6429292tendburg Ph.D., Carol
 REGISTRATION NUMBER: 39,317
 REFERENCE/DOCKET NUMBER: 190106,404
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 602 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 5:
 US-09-151-957-5

Query Match 49.6%; Score 1664.5; DB 4; Length 602;
 Best Local Similarity 54.0%; Pred. No. 8,2e-130;
 Matches 324; Conservative 93; Mismatches 166; Indels 17; Gaps 10;

QY 40 VRPQRTSSRELVLNLDGLMKFAL--ASGLNDTAQPTAPLPKGLCECPVSPASYNDITISRE 96
 DB 2 LRREVTPTREIKKLDELMAFSLDRENGCIDQ--RWMESALQESRAIAVGSFNDQFADAD 59
 QY 97 IHDHVGVVYQREIVYIPKMSQERIVLRASATHHGRIVYNNRLVAEHVGGYTFPEADYT 156
 DB 60 IRNYAGVWVYQREIVYIPKMSQERIVLRASATHHGRIVYNNRLVAEHVGGYTFPEADYT 119
 QY 157 ELVAPEKEFLITGVNNELTHEITIPPKITITGNATGKRIQTYOHDFYNYAGLARSIWLYS 216
 DB 120 PYVIAGKSVRITVCVNNELMWOTIPPGWVIT--DENKKKQSYHDFPNYAGIRSWLYT 178
 QY 217 VPQOHQIDITVTVTV--DGDNGLINYEVEVANOQTGOIQISVIDEGAIYAKASGAQGY 274
 DB 179 TPRTVVDITVTVTVHVAQDCNHAASVDWQV--VAN---GDVSEVELDADQOVVATGQTSGLT 234

QY 275 TIPSVKLMQPGAAYLYOLOVNIWSSGDVVDYTNLATGVRTVYVAGSQFLINKPEPYFTG 334
 DB 235 QVNPFLHMQPGEYLYELCY--TAKSQTECDIYPLAVGRSAVAKGEQFLINKPEPYFTG 292
 QY 335 FGRHEDTAVRGKGDPAVYVHDFOLMKWIGANSFRTSHYPYAEVMDPADRNGIIVIDET 394
 DB 293 FGRHEDADIRGKGFVDVYLWHDALMDWIGANSFRTSHYPYAEVMDPADRNGIIVIDET 352
 QY 395 PAVGNIAL--MGVSESGAP--QFTTPAINDKTOEAKHAKIRLADKXHASVYVMSIAN 452
 DB 353 AAVGNLSLIGFEAGNKKRELVEEAVNGEQAHQAIKELIADKXHASVYVMSIAN 412
 QY 453 EPASHEDGAREYEPFLNLTROLDPRPTFPAVYATVQOLDRISLDFVSCINRYFGWY 512
 DB 413 EPPTROQ--VHGNISPLAEATRKLDPRPTTCVNVMCDAHTDITISLDFVCLNRYFGWY 471
 QY 513 SQTGDEAEBALEKEHMOEKFRPIVWTEYGADTLAHLISLGLPMSSEFOVQMLDM 572
 DB 472 VQSGDELEAKVLEKELAMQEKHQPIITTEYGVDTLGLHSMYTDMSSEYQCAMLDM 531
 QY 573 YHRVPRISMAGEHYWNPADPOTNIGIIRVDGNKKGVFTDRKPKAAHSLPARFTSID 632
 DB 532 YHRVPRISAVAGEQVWNPADPOTNIGIIRVDGNKKGVFTDRKPKAAHSLPARFTSID 591

RESULT 12

5432081-2
 ; Patent No. 5432081
 ; APPLICANT: JEFFERSON, RICHARD A.
 ; TITLE OF INVENTION: HOST CELLS TRANSFORMED WITH THE E. COLI
 ; GLUCURONIDE PERMEASE GENE
 ; NUMBER OF SEQUENCES: 10
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/138,546
 ; FILING DATE: 15-OCT-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 447,976
 ; FILING DATE: 08-DEC-1989
 ; APPLICATION NUMBER: 264,586
 ; FILING DATE: 31-OCT-1988
 ; APPLICATION NUMBER: 119,102
 ; FILING DATE: 10-NOV-1987
 ; SEQ ID NO: 2:
 ; LENGTH: 602
 ; 5432081-2

Query Match 49.6%; Score 1664.5; DB 6; Length 602;
 Best Local Similarity 54.0%; Pred. No. 8,2e-130;
 Matches 324; Conservative 93; Mismatches 166; Indels 17; Gaps 10;

QY 40 VRPQRTSSRELVLNLDGLMKFAL--ASGLNDTAQPTAPLPKGLCECPVSPASYNDITISRE 96
 DB 2 LRREVTPTREIKKLDELMAFSLDRENGCIDQ--RWMESALQESRAIAVGSFNDQFADAD 59
 QY 97 IHDHVGVVYQREIVYIPKMSQERIVLRASATHHGRIVYNNRLVAEHVGGYTFPEADYT 156
 DB 60 IRNYAGVWVYQREIVYIPKMSQERIVLRASATHHGRIVYNNRLVAEHVGGYTFPEADYT 119
 QY 157 ELVAPEKEFLITGVNNELTHEITIPPKITITGNATGKRIQTYOHDFYNYAGLARSIWLYS 216
 DB 120 PYVIAGKSVRITVCVNNELMWOTIPPGWVIT--DENKKKQSYHDFPNYAGIRSWLYT 178
 QY 217 VPQOHQIDITVTVTV--DGDNGLINYEVEVANOQTGOIQISVIDEGAIYAKASGAQGY 274
 DB 179 TPRTVVDITVTVTVHVAQDCNHAASVDWQV--VAN---GDVSEVELDADQOVVATGQTSGLT 234
 QY 275 TIPSVKLMQPGAAYLYOLOVNIWSSGDVVDYTNLATGVRTVYVAGSQFLINKPEPYFTG 334
 DB 235 QVNPFLHMQPGEYLYELCY--TAKSQTECDIYPLAVGRSAVAKGEQFLINKPEPYFTG 292
 QY 335 FGRHEDTAVRGKGDPAVYVHDFOLMKWIGANSFRTSHYPYAEVMDPADRNGIIVIDET 394
 DB 293 FGRHEDADIRGKGFVDVYLWHDALMDWIGANSFRTSHYPYAEVMDPADRNGIIVIDET 352

Db 293 FGRHEDADLRGKGFQVNLVMDHALMDWIGANSYRSHYPAEMLDWADEHGIIVIDET 352
 QY 395 PAVGLNIAL-MGVSESGAP-QTFPPDAINDKTOEAKHQAIRELIARDKNHASVVMWSIAN 452
 Db 353 AAVGFNLISLIGIFGFAAGKPKELVSEBAVNGETQOAHLOAIKELIARDKNHPSVVMWSIAN 412
 QY 453 EPASHEDGAREYEPFLTNLTROLDPTPRITFANVGATATYQOLDRIISLDFVSCINRYFGWY 512
 Db 413 EPDTRPQ-VHGNISPLAEATRKLDPTPRITCVNVMFCDAHTDTISDLDFVCLNRYGYM 471
 QY 513 SQTGDLLEAEALAEKELHGMQEKFRPIVWTEYGADTLAHSISLGLPMSSEFOVQWMDM 572
 Db 472 VQSGDLETAKEVLEKELIWMQEKIHOPIITEYGVDTLAGHSMTYDMSSEYQCAWMDM 531
 QY 573 YHVFPRISMAHEWNNFADPOTNIGIIRVDGNKKGVFTDRDRPKRAAHSIRARWTSID 632
 Db 532 YHVFDRVSAVGEQVWNNFADPOTNIGIIRVDGNKKGVFTDRDRPKRAAHSIRARWTSID 591

RESULT 13

5432081-2
 Patent No. 5432081
 APPLICANT: JEFFERSON, RICHARD A.
 TITLE OF INVENTION: HOST CELLS TRANSFORMED WITH THE E. COLI
 GLUCONONIDE PERMEASE GENE
 NUMBER OF SEQUENCES: 10
 CURRENT APPLICATION NUMBER: US/08/138,546
 FILING DATE: 15-OCT-1993
 PRIOR APPLICATION NUMBER: 447,976
 FILING DATE: 08-DEC-1989
 APPLICATION NUMBER: 264,586
 FILING DATE: 31-OCT-1988
 APPLICATION NUMBER: 119,102
 FILING DATE: 10-NOV-1987
 SEQ ID NO: 2
 LENGTH: 602

Query Match 49.6%; Score 1664.5; DB 6; Length 602;
 Best Local Similarity 54.0%; Pred. No. 8.2e-130;
 Matches 344; Conservative 93; Mismatches 166; Indels 17; Gaps 10;
 QY 40 VBPQRTSSRELVLNDGLMKFAL--ASGLNDPAQWTPAPLPKGLCECPVASYNDIFTSRE 96
 Db 2 LRVEPTREIKKLDGLMAFSLDRENCIDQ--RWMSALQESRAIAPVGSFNQFADAD 59
 QY 97 IDHVGWVYVYOREVIVPKGMSOERYLVRAESATHGRIYVNNRLVAEHVGITPFEADV 156
 Db 60 IRNYAGNVVYVYOREVIVPKGMSOERYLVRAESATHGRIYVNNRLVAEHVGITPFEADV 119
 QY 157 ELVAPGEKRLITGVNNELTHEITPPKITTGNATGKRIQYQHDFNYVAGLARSIMLYS 216
 Db 120 PYVIAGKSVRIYVCNNELMQITIPPGWVIT--DENGKKKQSYFHFENYAGIHSVMLYT 178
 QY 217 VPQOHIODITVYTDV--DGNGLINYEVEVANQTTGOIQISVIDEGAIYAKASAGQTV 274
 Db 179 TPNTWDDITVTVHVADCNHASVDWQV-VAN--GDVSELRLADQOVVATGQTSGLT 234
 QY 275 TIPSVKLMOPGAAYLYOLOVNIYVSSGDVVDYTNLATGVRYTVKVASQFLINGKPEYFTG 334
 Db 235 QVAVPHLMOPGEGYIVELCV--TAKSQTECDIYPLRAGIRSAVVGEOFLINHKPEYFTG 292
 QY 335 FGRHEDTAVNGKGDPAVYVNDFOVMKWTIGANSFRTSHYPAEEMDWADEHGIIVIDET 394
 Db 293 FGRHEDADLRGKGFQVNLVMDHALMDWIGANSYRSHYPAEEMDWADEHGIIVIDET 352
 QY 395 PAVGLNIAL-MGVSESGAP-QTFPPDAINDKTOEAKHQAIRELIARDKNHASVVMWSIAN 452
 Db 353 AAVGFNLISLIGIFGFAAGKPKELVSEBAVNGETQOAHLOAIKELIARDKNHPSVVMWSIAN 412
 QY 453 EPASHEDGAREYEPFLTNLTROLDPTPRITFANVGATATYQOLDRIISLDFVSCINRYFGWY 512

Db 413 EPDTRPQ-VHGNISPLAEATRKLDPTPRITCVNVMFCDAHTDTISDLDFVCLNRYGYM 471
 QY 513 SQTGDLLEAEALAEKELHGMQEKFRPIVWTEYGADTLAHSISLGLPMSSEFOVQWMDM 572
 Db 472 VQSGDLETAKEVLEKELIWMQEKIHOPIITEYGVDTLAGHSMTYDMSSEYQCAWMDM 531
 QY 573 YHVFPRISMAHEWNNFADPOTNIGIIRVDGNKKGVFTDRDRPKRAAHSIRARWTSID 632
 Db 532 YHVFDRVSAVGEQVWNNFADPOTNIGIIRVDGNKKGVFTDRDRPKRAAHSIRARWTSID 591

RESULT 14

US-09-488-270A-2
 Sequence 2, Application US/09488270A
 Patent No. 6455759
 GENERAL INFORMATION:
 APPLICANT: Vilestra, Richard D
 APPLICANT: Walker, Joseph M
 TITLE OF INVENTION: Production of Multiple Proteins in Plants
 FILE REFERENCE: 960296.96501
 CURRENT APPLICATION NUMBER: US/09/488,270A
 NUMBER OF SEQ ID NOS: 4
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 2
 LENGTH: 1242
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE: Description of Artificial Sequence: Fusion Protein
 OTHER INFORMATION: Expression Cassette
 US-09-488-270A-2

Query Match 49.6%; Score 1664.5; DB 4; Length 1242;
 Best Local Similarity 54.0%; Pred. No. 2.5e-129;
 Matches 324; Conservative 93; Mismatches 166; Indels 17; Gaps 10;
 QY 40 VBPQRTSSRELVLNDGLMKFAL--ASGLNDPAQWTPAPLPKGLCECPVASYNDIFTSRE 96
 Db 642 LRVEPTREIKKLDGLMAFSLDRENCIDQ--RWMSALQESRAIAPVGSFNQFADAD 699
 QY 97 IDHVGWVYVYOREVIVPKGMSOERYLVRAESATHGRIYVNNRLVAEHVGITPFEADV 156
 Db 700 IRNYAGNVVYVYOREVIVPKGMSOERYLVRAESATHGRIYVNNRLVAEHVGITPFEADV 759
 QY 157 ELVAPGEKRLITGVNNELTHEITPPKITTGNATGKRIQYQHDFNYVAGLARSIMLYS 216
 Db 760 PYVIAGKSVRIYVCNNELMQITIPPGWVIT--DENGKKKQSYFHFENYAGIHSVMLYT 818
 QY 217 VPQOHIODITVYTDV--DGNGLINYEVEVANQTTGOIQISVIDEGAIYAKASAGQTV 274
 Db 819 TPNTWDDITVTVHVADCNHASVDWQV-VAN--GDVSELRLADQOVVATGQTSGLT 874
 QY 275 TIPSVKLMOPGAAYLYOLOVNIYVSSGDVVDYTNLATGVRYTVKVASQFLINGKPEYFTG 334
 Db 875 QVAVPHLMOPGEGYIVELCV--TAKSQTECDIYPLRAGIRSAVVGEOFLINHKPEYFTG 932
 QY 335 FGRHEDTAVNGKGDPAVYVNDFOVMKWTIGANSFRTSHYPAEEMDWADEHGIIVIDET 394
 Db 933 FGRHEDADLRGKGFQVNLVMDHALMDWIGANSYRSHYPAEEMDWADEHGIIVIDET 352
 QY 395 PAVGLNIAL-MGVSESGAP-QTFPPDAINDKTOEAKHQAIRELIARDKNHASVVMWSIAN 452
 Db 993 AAVGFNLISLIGIFGFAAGKPKELVSEBAVNGETQOAHLOAIKELIARDKNHPSVVMWSIAN 412
 QY 453 EPASHEDGAREYEPFLTNLTROLDPTPRITFANVGATATYQOLDRIISLDFVSCINRYFGWY 512
 Db 1093 EPDTRPQ-VHGNISPLAEATRKLDPTPRITCVNVMFCDAHTDTISDLDFVCLNRYGYM 1111
 QY 513 SQTGDLLEAEALAEKELHGMQEKFRPIVWTEYGADTLAHSISLGLPMSSEFOVQWMDM 572
 Db 1112 VQSGDLETAKEVLEKELIWMQEKIHOPIITEYGVDTLAGHSMTYDMSSEYQCAWMDM 1171

Qy 573 YHRVFDRIESMAGEHYVNFADFTQNLGIRVDGNKKGVFTDRKPKAAASHLRARWTSID 632
| | | | | : : : | | | | | : : : | | | | | : : : | | | : : :
Db 1172 YHRVFDRIESMAGEHYVNFADFTQNLGIRVDGNKKGVFTDRKPKAAASHLRARWTSID 1231

RESULT 15
5268463-2
; Patent No. 5268463
; APPLICANT: JEFFERSON, RICHARD A.
; TITLE OF INVENTION: PLANT PROMOTER a-GLUCURONIDASE GENE
; CONSTRUCT
; NUMBER OF SEQUENCES: 9
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/447,976
; FILING DATE: 08-DEC-1989
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 119,102
; FILING DATE: 10-NOV-1987
; APPLICATION NUMBER: 264,586
; FILING DATE: 31-OCT-1988
; SEQ ID NO: 2:
; LENGTH: 600
5268463-2

Query Match 48.9%; Score 1638.5; DB 6; Length 600;
Best Local Similarity 53.8%; Pred. No. 1,2e-127;
Matches 333; Conservative 93; Mismatches 165; Indels 19; Gaps 12;

Qy 40 VREPORTSSRELVLNDGLMKFAL--ASGLNDTAQPTAPLPKGLCCPVASYNDFISRE 96
| | | | | : : : | | | | | : : : | | | | | : : : | | | : : :
Db 2 LRPEVPTREIKLDGLMAFSLDRNGCIDQ--RWESALQESRAIAYPSFNDQFADAD 59
Qy 97 IHDVGVWVYQREVIYPKGSQERYLVRAESATHHGRIVYNNRLVAEHVGGTTFEADYT 156
| | | | | : : : | | | | | : : : | | | | | : : : | | | : : :
Db 60 IRNYAGNVWYQREVIYPKGSQERYLVRAESATHHGRIVYNNRLVAEHVGGTTFEADYT 119
Qy 157 ELVAPGEKRLTGVNNELTHETIPPKKITTGATGRIQIOTYQHPYNAVGLARSIMLYS 216
| | | | | : : : | | | | | : : : | | | | | : : : | | | : : :
Db 120 PYVIAGKSVRIIVCANNELMWOTIPPGMYTT--DENGKKQSYFHDFFNAGIHRSMWLYT 178
Qy 217 VPQOHIDITVTVDV--DGDNGLIYNEVEVANQTTGQIQISVIDEDGATVAKASGAQTV 274
| | | | | : : : | | | | | : : : | | | | | : : : | | | : : :
Db 179 TPRTWVDITVTVHVAQDCNHAQSVDMQV--VAN---GDVSELRDADQVAVATGQTSGLT 234
Qy 275 TIPSVKLQPGAAVLYQLOVNVSSGDVVDTYNLATGVRTVYVAGSQFLNGKPPYFTG 334
| | | | | : : : | | | | | : : : | | | | | : : : | | | : : :
Db 235 QVNPPLMOPG--EYLVELCV--TAKSQTECDIPLRAGIRSVAVKGEQFLNHKPEYFTG 291
Qy 335 FGHEDTAVRGKGDPAVWYHDQOLMKWTIGANSFRTSHYPAEYVDFADRNQIVVIDET 394
| | | | | : : : | | | | | : : : | | | | | : : : | | | : : :
Db 292 FGHEDTAVRGKGDPAVWYHDQOLMKWTIGANSFRTSHYPAEYVDFADRNQIVVIDET 351
Qy 395 PAVGLNIAL--MGVSESGAP--QFTPDALNDKTOEAKHQAIRELIARDKNHASVVMMSIAN 452
| | | | | : : : | | | | | : : : | | | | | : : : | | | : : :
Db 352 AAVGFNLSTGIGFEAGNKREKELSEBAVNGETQAHQAKEIARDKNHPSVVMMSIAN 411
Qy 453 EPASHEDGAREYFEPLTNLTROLDPTRPITFEANVGTATYQLDRIISDLFDVSCINRYFGWY 512
| | | | | : : : | | | | | : : : | | | | | : : : | | | : : :
Db 412 EPDTRPQ--VHGNISPLAEATRKLDPTRPITCVNMFCDATDTISDLFDVLCINRYFGWY 470
Qy 513 SQTGDLFEAALAEKELHGOEKFHRPIVWTEGADTLAGIHSILGLPWSEEFQVOMLDM 572
| | | | | : : : | | | | | : : : | | | | | : : : | | | : : :
Db 471 VOSGDLETAEKELAEKELHGOEKFHRPIVWTEGADTLAGIHSILGLPWSEEFQVOMLDM 530
Qy 573 YHRVFDRIESMAGEHYVNFADFTQNLGIRVDGNKKGVFTDRKPKAAASHLRARWTSID 632
| | | | | : : : | | | | | : : : | | | | | : : : | | | : : :
Db 531 YHRVFDRIESMAGEHYVNFADFTQNLGIRVDGNKKGVFTDRKPKAAASHLRARWTSID 589

Search completed: March 18, 2005, 23:53:55
Job time : 46 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: March 18, 2005, 23:33:12 ; Search time 171 Seconds
(without alignments)
1433.954 Million cell updates/sec

Title: US-10-757-093-4
Perfect score: 3354
Sequence: 1 MKFLTGLSLSLAAPSIGTP.....RKPKAAHSLRAWTSIDKN 634

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues
Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1980s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1756	52.4	602	3	AA828409
2	1714.5	51.1	832	2	AAW04302
3	1714	51.1	618	7	ADL01666
4	1714	51.1	1010	3	AA868840
5	1711	51.0	603	2	AAW93827
6	1711	51.0	603	2	AAW93824
7	1711	51.0	603	2	AA828431
8	1711	51.0	603	5	AB884108
9	1711	51.0	603	6	ABP96657
10	1711	51.0	603	6	ABU19649
11	1711	51.0	603	8	ADN18189
12	1706	50.9	604	7	ADD27986
13	1706	50.9	659	7	ADD27989
14	1706	50.9	850	7	ADD27991
15	1697	50.6	603	5	AB884107
16	1667.5	49.7	602	2	AA84387
17	1665.5	49.6	607	8	AD826258
18	1665	49.6	711	6	AB883626
19	1664.5	49.6	602	2	AAW42429
20	1664.5	49.6	1242	5	AB884108
21	1664.5	49.6	1242	6	AB884637
22	1657.5	49.4	602	1	AA82948
23	1590.5	47.4	617	4	AAU39683
24	1590.5	47.4	617	6	ABM36202
25	1439	42.9	615	3	AA828408

26	1439	42.9	618	2	AAW93821	AAW93821	Bacillus
27	1437	42.8	602	2	AAW93825	AAW93825	Bacillus
28	1437	42.8	602	2	AAW93822	AAW93822	Bacillus
29	1437	42.8	602	2	AAW93826	AAW93826	Bacillus
30	1437	42.8	602	3	AAW93820	AAW93820	Staphylococcus
31	1431	42.7	602	2	AAW93823	AAW93823	Bacillus
32	1338	39.9	613	2	AAW93828	AAW93828	E. coli G
33	1338	39.9	613	3	AAW93828	AAW93828	Human GUS
34	1338	39.9	613	3	AAW93828	AAW93828	Escherich
35	1338	39.9	613	4	AAW93828	AAW93828	Mutant he
36	1338	39.9	613	4	AAW93828	AAW93828	Heavy cha
37	1338	39.9	613	4	AAW93828	AAW93828	Human bet
38	1338	39.9	613	7	AAW93828	AAW93828	Human pro
39	1338	39.9	613	7	AAW93828	AAW93828	Human pro
40	1338	39.9	613	8	AAW93828	AAW93828	Protein e
41	1338	39.9	613	8	AAW93828	AAW93828	Antagonis
42	1338	39.9	722	6	AAW93828	AAW93828	L. mexica
43	1338	39.9	722	6	AAW93828	AAW93828	Human bet
44	1338	39.9	722	8	AAW93828	AAW93828	Fusion pr
45	1335	39.8	648	7	AAW93828	AAW93828	Rat Prote

ALIGNMENTS

RESULT 1
AAB28409
ID AAB28409 standard; protein; 602 AA.
XX
AC AAB28409;
XX
DT 26-JUN-2001 (first entry)
XX
DE Salmonella beta-glucuronidase.
XX
KM Microbial; beta-glucuronidase; GUS; Enterobacter; Salmonella;
KW Pseudomonas; Staphylococcus; Thermotoga; transgenic plant; bioindicator;
KW transgenic insect; marker; glucuronide detoxification.
XX
OS Salmonella sp.
XX
PN WO200055333-A1.
XX
PD 21-SEP-2000.
XX
PF 16-MAR-2000; 2000WO-US007107.
XX
PR 17-MAR-1999; 99US-00270957.
XX
PA (CAMP-) CAMBIA BIOSYSTEMS LLC.
XX
PI Jefferson RA, Mayer JE;
XX
DR WPI: 2000-647075/62.
XX
DR N-PSDB; AAA07939.
XX
PT Novel microbial beta-glucuronidase genes and gene products used as
PT reporter/effector molecule, as diagnostic tool, in positive selection, to
PT target molecules to specific cells and to detect and track linked genes.
XX
PS Example 3; Fig 17; 116pp; English.
XX
CC The present sequence is a microbial beta-glucuronidase (GUS) protein. GUS
CC genes were obtained from six different genera: Enterobacter/Salmonella,
CC Pseudomonas, Salmonella, Staphylococcus and Thermotoga. Microbial GUS can
CC be used as a reporter/effector molecule for transgenic constructions and
CC in vitro diagnostic applications. It may also be used to generate
CC sentinel plants that serve as bioindicators of environmental status. It
CC may be used to generate transgenic insects for tracking insect
CC populations or to facilitate the development of a bioassay for compounds
CC that affect molecules critical for insect development (e.g. juvenile
CC hormone). Secreted GUS may also serve as a marker for beneficial fungi
CC destined for release into the environment. In animal systems, secreted

CC GUS may be used to achieve extracellular detoxification of glucuronides
 CC (e.g. toxin glucuronide) and to examine conjugation patterns of
 CC glucuronides. Microbial GUS may also be used in traditional medical
 CC diagnostic assays, for drug testing, pharmacokinetic studies,
 CC bioavailability studies, diagnosis of diseases and syndromes, following
 CC progression of disease or its response to therapy. Microbial GUS has
 CC increased thermal stability, high turnover number and enzymatic activity.
 CC It is highly specific for the substrate and water soluble, and the
 CC substrates are stable

XX Sequence 602 AA;

Query Match 52.4%; Score 1756; DB 3; Length 602;

Best Local Similarity 55.1%; Pred. No. 3.8e-134;

Matches 329; Conservative 99; Mismatches 159; Indels 10; Gaps 6;

QY 40 VRPQRSSRLVNLDDGKFKALASGLNDTRP-WTAPLPGLCEPPASINDIFISREH 98
 DB 2 LRSVETATREIKLIDGMSFCMDESECGNAQWMPRLPQSRALAVPGSYNDQFAAEIR 61
 QY 99 DHGVWVYQREVIIVPKGMSQERYLVRAESATHHGRIVYNNRLVAEHVGYTPPEADVTTEL 158
 DB 62 NYGVNWWYQREIRIPKMDRQIVLRFDATVHGKAVNDQFLMEHQGRTTPEADISHL 121
 QY 159 VAPGEKRLTGVNNELTETTPGKITTNATGKRIQYQHPYVAGIARSIMLYSV 218
 DB 122 ISAGESVRIIVCVNNELMTQITPPGVVYQG-VNGKKQAFYHFFVYAGIHSVIMLYTTP 180
 QY 219 QGHIDITVTVDDGNGNLNVEVANQTTQIQISVIDEDDAIYAKSAGCTYIIS 278
 DB 181 KTRVEDITVTVQVADD-LAQATVAMOVNANGVRELRABEQQLVASGGEGKGLLEGG 238
 QY 279 VKLMQGAAYLYOLQVNIYVSSGDVDTYNLATGVRVTVKVASQFLINGKPYFTGFGKH 338
 DB 239 PRLMOGEGVLYELRV-IAQHQBDEDEYELRGISVEVKGQFLINHPFTFTGGRH 296
 QY 339 EDPAVNGKGDPAVYVNDFOVMKMGANSFRITSHYPAEEMDPADRNGLIVIDEFPVAG 398
 DB 297 EDADLKGKGDVNLVMDHMLDMIGANSYRISHYPAEEMDLMADEHGIVIIDETPAVVG 356
 QY 399 LNTALNGVS--ESGAPQPTFPAINDKTOEAKHQAIRELIARDKNHASVVMSSIANEPA 455
 DB 357 FNLSTL-GISDVGEKREKELYSDEAVNDETORALQAIKELIARDKNHPSVVMSSIANEPA 415
 QY 456 SHEDGAREYFEPLNLTROLPTPTPTTFANVATATYQDLRISDLPVSCINRYFGWYSQT 515
 DB 416 TRNGAREYFAPLAQTRLEDPTRPITCVVMFCDAESDTITDLPVVCINRYGYVQS 475
 QY 516 GDLEEAALAEKELHGMQEKFRPIYVTERGADTLAHLISLGLPWSEEPQVOMLDMYHR 575
 DB 476 GDLEKAKEVLEKELAMOEKLRPIITTEGYVDTLAGLSHMYNDMWSSEEQCAMLDMYHR 535
 QY 576 VFPRISSMAGEHWNPADFOYNLGIIRVQDNKGVFTPRDKPKAAHSLPARWTSID 632
 DB 536 VFPRISSAVVGEQWNPADFPATISQINRVGANKGIFTRDKPKASAPFLQKWTGMD 592

RESULT 2
 ID AAM04302
 ID AAM04302 standard; protein; 832 AA.
 AC AAM04302;
 XX 25-MAR-2003 (revised)
 DT 16-FEB-1997 (first entry)
 XX Antibody/beta glucuronidase fusion protein.
 DE Antibody; fusion protein; recombinant antibody; tumour therapy; produg.
 KW Synthetic.
 OS EP737747-A2.

XX 16-OCT-1996.
 PD 13-MAR-1996; 96EP-00103913.
 PF 11-APR-1995; 95DE-01013676.
 PR (BEHN) BEHRINGER AG.
 PA Oper M, Bosset K, Czech J;
 XX WPI; 1996-457328/46.
 DR N-PSDB; AAT18397.
 XX Prodn. of recombinant antibody (Ab), Ab fragment or Ab/enzyme fusion
 PT protein - by cytoplasmic expression in thio:redoxin:reductase deficient
 PT E. coli.
 PS Example 1; Fig 5a; 12pp; German.

CC Production of recombinant antibodies (Ab), Ab fragments or Ab
 CC fragment/enzyme fusion proteins can be used for tumour therapy,
 CC especially when the fusion protein comprises a tumour-specific Ab
 CC fragment and an enzyme capable of converting a non-toxic produg to a
 CC toxic drug. The fusion proteins are constructed in expression vectors and
 CC expressed in thio:redoxin:reductase deficient E.coli, allowing expression
 CC products to be isolated in soluble functional form without renaturation.
 CC The Ab fragment is an Fab fragment or an antigen binding region. In the
 CC fusion protein, the Ab component is humanised and the enzyme component is
 CC a human cytoplasmic enzyme. This fusion protein comprises the antibody
 CC constant and variable heavy chain regions and the E.coli Beta
 CC glucuronidase enzyme. (Updated on 25-MAR-2003 to correct PR field.)

XX Sequence 832 AA;

Query Match 51.1%; Score 1714.5; DB 2; Length 832;

Best Local Similarity 52.7%; Pred. No. 1.5e-130;

Matches 338; Conservative 98; Mismatches 180; Indels 25; Gaps 11;

QY 8 SLTSLAAPSGLTPA-----ARHPRNEMTQHEQPL-----IKVRPQRTSSRLVNLGKLMK 58
 DB 190 SVVTVPPSSSLGTQYTCNNVHKRPSTVKDKRYELSSGGSVMKRPVETTPPEIKKLDGLMA 249
 QY 59 PAL--ASGLNDTAQPTAPLPKGLCEPPASVNDIFISREIHADVGVWYQREVIIVPKG 115
 DB 250 FSLDENCGIDQ--RWMSALQESRALAVPGSFNDQFADIRYAGVNWYQREVIIVPKG 307
 QY 116 WQGERLYVRAESATHHGRIVYNNRLVAEHVGYTPPEADVTTELVAPEKFRITGVNNEL 175
 DB 308 MAGORIVLRFDAVTHYGVKVVANNOEVMERHOGGYTPPEADVTPEYVLAGKSVRITVCNNEL 367
 QY 176 THEITPFGKITGNATGKRIQYQHPYVAGIARSIMLYSVPOQHIDITVTVV--DG 233
 DB 368 NMQITPFGKIT--DENGKKKQSYFNFYPAISHSVMLYTFENWVDDITVTVHAQDC 426
 QY 234 DNGLINYEVEVANQTTQIQISVIDEDGAIYAKSAGACTVTTPSVKLMQGAAYLYOLO 293
 DB 427 NHASVDQW-VAN--GDVSVELRDADQGVATGGTSGTLQVNPMLMQPSEGVLYELC 482
 QY 294 VNIYSSGDVDTYNLATGVRVTVKVASQFLINGKRPFTFGGKEDPAVNGKGDPAVM 353
 DB 483 V--TAKSQTECDIYPLRGIRSVAAVKGEOFLINHPKFFFTGGRHEDADLRKKGFDNYLM 540
 QY 354 VHDPLMKMGANSFRITSHYPAEEMDPADBNGLVIVDEFPVAGLNTAL-MGVSBSGAP 412
 DB 541 VHDHALMDWIGANSFRITSHYPAEEMDLADDEHGIVVIDETRAVGFNLISLGIFGAGNKP 600
 QY 413 -QTFPDAINDKTOEAKHQAIRELIARDKNHASVVMSSIANEPASHEDGAREYFEPLNLT 471
 DB 601 KELYSEAVNGEITQAHQAIKELIARDKNHPSVVMSSIANEPDRPQAGAREYFAPLAEA 660
 QY 472 TRQDPTPTPTTFANVATATYQDLRISDLPVSCINRYGYWYSQTDLEAEALAEKELHGM 531

DB 661 TRKLDTLPRTTCVNVNMFCDNADHTDITSDLPVLCINRYYGWYVSGDLETAKEVLEKELLA 720
 QY 532 MOKFPRPIWTEYAGDTLAGLSIIGLPKSEERQVOMLMDYHRVFPDRISMAGEHVMNF 591
 DB 721 MOKLHQPITTEYGVDTLAGLSMTYDMWSEEQCAMLMDYHRVFPDRSAVVGEOVMNF 780
 QY 592 ADEFQNIIGIRVDGNKKGVFTDRDKPKYAAHSLRARWTSID 632
 DB 781 ADFATSGILRVGNGKKGIFTRDRKPKSAFLQKRTGMN 821

RESULT 3
 ADL01666
 ID ADL01666 standard; protein; 618 AA.
 XX

AC ADL01666;
 XX
 DT 06-MAY-2004 (first entry)
 XX

DE Modified GUS protein HGUH amino acid sequence.

KM GUS; beta-glucuronidase; fusion polypeptide; Extn; Extc; Int;
 KW N-terminal portion; Intein; C-terminal portion; plant optimised codon;
 KM protein splicing mechanism; transgene;
 XX multi-functional hybrid protein polymer; circular protein.

OS Unidentified.
 OS Synthetic.

PN WO2003066861-A1.

PD 14-AUG-2003.

PF 04-FEB-2003; 2003WO-US003435.

PR 04-FEB-2002; 2002US-0354395P.

PA (DUPO) DU PONT DE NEMOURS & CO E I.

PI Yadav NS, Yang JG;

XX WPI; 2003-731504/69.

PT New isolated polynucleotide encoding a polypeptide comprising an Extn, an
 PT Extc, and an Int interspersed between the Extn and the Extc, useful for
 PT Intein-mediated protein splicing.

PS Example 2; SEQ ID NO 28; 63pp; English.

CC This invention relates to a novel isolated polynucleotide comprising a
 CC nucleotide sequence that encodes a polypeptide or a fusion polypeptide
 CC comprising an Extn, an Extc and an Int interspersed between the Extn and
 CC the Extc. The Extn is the N-terminal portion of the polypeptide, the Int
 CC is an Intein, and the Extc is the C-terminal portion of the polypeptide.
 CC At least a portion of the nucleotide sequence has been modified to
 CC contain plant optimised codons. The polynucleotide and methods are useful
 CC in introducing a protein splicing mechanism into plants by employing
 CC inteins and transgenes. This permits in vivo and in vitro synthesis of
 CC homogeneous and large multi-functional hybrid protein polymers and
 CC circular proteins. The present sequence is that of a protein which is
 CC related to the invention.

XX
 SQ Sequence 618 AA;

Query Match 51.1%; Score 1714; DB 7; Length 618;
 Best Local Similarity 55.2%; Pred. No. 1.1e-130;
 Matches 331; Conservative 92; Mismatches 161; Indels 16; Gaps 9;

QY 40 VRQQRSSRELVLNDGIMKFAAL--ASGLNDTAQPTAPLPKGLGCVVPASYNDFISRE 96
 DB 9 VRPEPTTRBIKLDGIMAFSLDRENGIDO--RWMSALQESRAVLPDSFNDQFADAD 66
 QY 97 IHDHVGWVYQREVIYVKGMSQERVLVRAESATHHGRIVYNNRLVAEHVGQYTPFEADVT 156

DB 67 IRNVAGVWVYQREVFIPKMGAGQRIYLRFDAVTHYGKVVWNNQBNVHEHQGYTPFEADVT 126
 QY 157 ELVAPGEKRLTIGVNNELTHEITPPGKITTGATGRIOTYOHDFPNVAGLARSLTMS 216
 DB 127 PYVIAQSVARITVCVNNELMWQITPPGMVLT--DENGKKOSYFHDFFNVAGIRSVMLYT 185
 QY 217 VPOOHODIIVVTVDV--DQNGGLINYEVEVANAOTGQIQISVIDEDGAIYAKASGQSTV 274
 DB 186 TPTWYDDITVTVHVAQDCNHAQVDMQV--VAN--GDVSELADADQGVVATQGSIGTL 241
 QY 275 TIPSVKLMQFGAAYLYQLOQVNIWSSGDVVDTYNLATGVRTVYVAGSQFLINGKPFYFTG 334
 DB 242 QVNVPHLMQEGEGYLYELCV--TAKSQTECDIYPLRVGIRSVAVKGEQFLINHKPFYFTG 299
 QY 335 FGGHEDTAVRGKHDPAAYVHDFQLMKVIKIGANSFRTSHYVYAEVWDFARNQIIVTDET 394
 DB 300 FGRHEDADLRGKGFVDVVLVHDALMDWIGANSYRSHYVYAEVWMADEHGIIVTDET 359
 QY 395 PAVGNIAL-MGVSESGAP--QTFTPPAINDKTOEAKHQAIRELIARDKNHASYVMMSIAN 452
 DB 360 AAVGNLSLIGIEAGNKKPKELYSEAVNGETQAHQAIKELIARDKNHPSVYMSIAN 419
 QY 453 EPASHDEGAREYEPITNLTRQDPTRPITFANVGTATYQLDRIIDLFDVSCINRYFGWY 512
 DB 420 EPDTRPGAREYFAPLAELATRKLDPTRPITCVNVNMFCDNADHTDITSDLPVLCINRYYGWY 479
 QY 513 SQTGDLSEAAALEKELHGWQEKFRPIVWTEYAGDTLAGLSIIGLPKSEERQVOMLMD 572
 DB 480 VQSGDLETAKEVLEKELHGWQEKHPIITTEYGVDTLAGLSMTYDMWSEEQCAMLMD 539
 QY 573 YHRVFPDRISMAGEHVMNFPADFOTNIGIIRVDGNKKGVFTDRDKPKYAAHSLRARWTSID 632
 DB 540 YHRVFPDRISAVVGEQVMNFPADFPATSGILRVGNGKKGIFTRDRKPKSAFLQKRTGMN 599

RESULT 4
 ID AAY68840
 AAY68840 standard; protein; 1010 AA.

AC AAY68840;

DT 16-MAY-2000 (first entry)

DE Fusion protein of the DNA-binding domain of NAP and GUS.

KM ZAP2; poly(ADP-ribose) polymerase; PARP; poly(ADP-ribose) transferase;
 KM programmed cell death; apoptosis; growth rate; stress; cold; pathogen;
 KM pest; drought; heat; fungi; nematode; seed-shatter.

OS Synthetic.
 OS Arabidopsis thaliana.

PN WO200004173-A1.

PD 27-JAN-2000.

PF 12-JUL-1999; 99WO-EP004940.

PR 17-JUL-1998; 98US-00118276.

PA (PLBZ) PLANT GENETIC SYSTEMS NV.

PI Babyichuk E, Kushnir S, De Block M;

XX WPI; 2000-182436/16.

PT Modulating cell death, growth and stress resistance in eukaryotes.

PT Specifically plants, used, e.g to impart fungus or nematode resistance.

PS Disclosure; Page 112-116; 126pp; English.

XX The present sequence represents a fusion protein of the DNA-binding

[illegible]

Dd		235	QVNVPHLMQPEGVLYELCV--TAKSQGECDIYPLRVGRISAVXKGEFLLHNHKEFYPTG	292
Oy	"	335	EKGEDTVARVGKHDPVAMVNDFOQMKTIGANSFRTSHYPAAEVMYPADNGIVVIDET	394
Dd		293	FGRHEDDLRNKGCPDNYMLMVHDHALMDWIGANSTYRSHYPAAEMLDMWADEHGIVVIDET	352
Oy		395	PAYGLNTALMGVESGSAP-OTFPDAILNDTOEAHQAIIBELLARDKNHNSVMMWSIAN	452
Dd		353	AAVFENSLSLGGFEAGNKPKELYSEAAVNGTQQHLCALIKELLARDKNHNSVMMWSIAN	412
Oy		453	EPASHEDGAREFEPLTNLTQLPRTPIFTFANYGTATVQADRISDLDFVCSCINREFGMY	512
Dd		413	EPDRTPGARGRYFAPLAEATRKLDPTRITLVNMWFCDAHNDITSIDLFDVICNRYYGYM	472
Oy		513	SQTGDLEBAEALALEKEHLGWQEKFRPIVTWTGYADTLAHSILGLFWSEEFQVOMLDM	572
Dd		473	VQSODLETFAEKVYLEKELLAWOEKLIHOPIITIEYGVDTLAGLHSMWTDWMSSEYOCCAMLDM	532
Oy		573	YHRVPDPIESAGEHWNPFADPOTNLGIRPDNGKKGVFTPRDKPKAAHSLRAWTSID	632
Dd		533	YHRVFDRAVSVAVGQVNWAFADFATSQGLIRVGNKKGIETDRKPKSAAPLLOKRMTGMN	592
RESULT 7				
AAB28431				
ID	AAB28431	standard; protein; 603 AA.		
XX				
AC	AAB28431;			
XX				
DT	26-JAN-2001	(first entry)		
De		Human beta-glucuronidase HGUS.		
XX				
KW	Microbial; beta-glucuronidase; GUS; Enterobacter; Salmonella;			
KW	Pseudomonas; Staphylococcus; Thermotoga; transgenic plant; bioindicator;			
KW	transgenic insect; marker; glucuronide detoxification.			
OS	Homo sapiens.			
XX				
PN	WO200055333-A1.			
PD	21-SEP-2000.			
XX				
PF	16-MAR-2000; 2000WO-US007107.			
XX				
PR	17-MAR-1999; 99US-00270957.			
XX				
PA	(CAMB-) CAMBIA BIOSYSTEMS LLC.			
XX				
P1	Jefferson RA, Mayer JE;			
XX				
DR	WPI; 2000-647075/62.			
XX				
PT	Novel microbial beta-glucuronidase genes and gene products used as			
PT	reporter/effector molecule, as diagnostic tool, in positive selection, to			
PT	target molecules to specific cells and to detect and track linked genes.			
PS	Example 4; Fig 5A; 116pp; English.			
XX				
CC	The present sequence is a microbial beta-glucuronidase (GUS) protein. GUS			
CC	genes were obtained from six different genera: Enterobacter/Salmonella,			
CC	Pseudomonas, Salmoneila, staphylococcus and Thermotoga. Microbial GUS can			
CC	be used as a reporter/effector molecule for transgenic constructions and			
CC	in vitro diagnostic applications. It may also be used to generate			
CC	sentinel plants that serve as bioindicators of environmental status. It			
CC	may be used to generate transgenic insects for tracking insect			
CC	populations or to facilitate the development of a biosay for compounds			
CC	that affect molecules critical for insect development (e.g. juvenile			
CC	hormone). Secreted GUS may also serve as a marker for beneficial fungi			
CC	destined for release into the environment. In animal systems, secreted			
CC	GUS may be used to achieve extracellular detoxification of glucuronides			
CC	(e.g. toxin glucuronide) and to examine conjugation patterns of			
CC	glucuronides.Microbial GUS may also be used in traditional medical			

[illegible]

Oy	217	VPGQHIQIOTVTVTDV--DGDGNLIVYEVEVNAQTGGIOISIVIDEQGLVAALASAGQGVV	274
Dd	179	TPTNVTVDITVTTHVAQDCNHAASVUMQV-VAN---GDVSELRADQQVVAIGQGTSGTL	234
Oy	275	TIPSVKLMQPGAAVLYLOLVNIVIGSSGDVDTYNLATGVRTYKVASGOPLNGKPEYFTG	334
Oy	235	QVNNPHLMQPGEGVLYELCV--TAKSQTECDIYPLRVGIRSAVKGEQFLNNKPFYFTG	292
Dd	293	FGRHEDADLRKGGPENVLMVHDHALMDWIGANSYRTSHPYABEEMLDMADEHGIIVIDET	352
Oy	395	PAYGNTNL--NGVSSSGAP-QTFETDADINDKQOEAKHKAIRLILARDKNHASVVMMSIAN	452
Dd	353	AAVGNLSLGLGFPEAGNPKPELYSEAAVGEYQAHLQIKELIARDKNHPSVVMMSIAN	412
Oy	453	EPASHEDAREYFEBPLTNLTRQLDPTRPITFANVGTATQOLDRISDLFPVSCINRYFGWY	512
Dd	413	EPDRPQARREYFAPLAETRLDPTRPITTCVNVFPCDAHTDITISLFPVLCLNRYGYW	472
Oy	513	SQTGDLFEAEALAEKELHGMQEKFRPIVMTREYGDITLAGHSILGLPMSEEFQVOMLDM	572
Dd	473	VQSGDLFEAEKYLEKELHGMQEKHQPILITREYGVDTLAGHSIMYDMMSSEYQCAMLDM	532
Oy	573	YHRVYDRLESMAAGEVNNFADPQTNLGIIRVNDGNKKGVTRTRRKPRAAAHSLPARVTSID	632
Dd	533	YHRVFDRYSAAVGEQVNNFADFATISQGLIRVGNKKGIPTRRKPKSAALFLQKRWGMN	592
RESULT 10			
ABJ19649			
ID	ABJ19649	standard; protein; 603 AA.	
AC	ABJ19649;		
DT	03-APR-2003	(first entry)	
DE	Artificial plant chromosome related beta-glucuronidase SEQ ID No 17.		
XX	Plant artificial chromosome; PAC; transgenic plant; vaccine;		
KW	blood factor; hepbicide; stress; agronomical; nutrient quality;		
KW	bacterial artificial chromosome; BAC; yeast artificial chromosome; YAC;		
KW	enzyme.		
XX	Escherichia coli.		
XX	WO200296923-A1.		
PN	05-DEC-2002.		
XX	30-MAY-2002; 2002WO-US017451.		
PF	30-MAY-2001; 2001US-0294687P.		
PR	04-JUN-2001; 2001US-02396329P.		
PA	(CHRO-) CHROMOS MOLECULAR SYSTEMS INC.		
PA	(AGRI-) AGRISOMA INC.		
XX	Perez C, Fabijanski SF, Perkins E;		
DR	WPI; 2003-140436/13.		
DR	N-PSDB; ABT16606.		
XX	Producing artificial chromosome by introducing a nucleic acid info plant		
PT	cell, selecting artificial chromosome that has one or more repeat regions		
PT	with equivalent amounts of euchromatic and heterochromatic nucleic acids.		
PS	Disclosure; Page 245-246; 269pp; English.		
CC	The invention relates to a novel method for producing plant artificial		
CC	chromosomes. The invention also relates to methods for targeting		
CC	insertion of heterologous DNA into plant artificial chromosomes, methods		
CC	for delivery of plant chromosomes to selected cells and tissues. The		

QY 157 ELVAPGKFLITGVNNELTHEITIPGKITGNATGRIQIOTYOHDFYVAGLSRWLS 216
CC The invention relates to a method of expressing a heterologous
CC polypeptide by a host cell. The methods are useful for producing altered
CC seed meals by manipulating oil bodies of plants. The present sequence is
CC used in the exemplification of the present invention.
XX
SQ Sequence 604 AA:
QY 275 TITSVVLMOGAAVLYOLOVNISSGSDVDVTYNLATGVTVVAVSQFLINKPFFETG 334
Db 235 QVNPFLMKGEGEYLYELCV--TAKSQTECDIPLRVGIRSVAVKGOFLINHPFFETG 292
QY 335 FGHEDTAVRGKGDPAVYVHDFOLMKMGIGANSFRISHYAEVMDPADRNGIVDET 394
Db 293 FGHEDADLRGKGFVDVYVHDHADMWIGANSFRISHYAEVMDPADRNGIVDET 352
QY 395 PAVGNIAL-MGVSESGAP-QTFPPDAINDKTOEAKHQAIRELIARDKNASVVMMSIAN 452
Db 353 AAVGFSLSLIGIPEAGNKPKEYSEAVNGETOAHQAIRELIARDKNHPSVVMMSIAN 412
QY 453 EPASHEDGAREYEPFLTNLTROLDPTFPITFANVGTATYOLDRIISDLFDVSCINRYGWY 512
Db 413 EPDTRPGAREYFAPLAETARKLDPTFPITCVMMFCDAHTDITSDLDVCLNRYGWY 472
QY 513 SQTGDLSEAAALEKELHGWQEKFRPIVMTVEGADTLAHLSTGLPMSEEPQVQWLDL 572
Db 473 VQSGDLETAKEVLEKELAMOEKIHOPITTEYGVDTLAGHSVYTDMSSEETQCAWLDL 532
QY 573 YHRVFDRIESMAGBHVNFADPOTNLGIIIRVDSNKGKGFTRDRKPKAAASLRARWTSID 632
Db 533 YHRVFDRIESVAVGEQVNFADPOTNLGIIIRVDSNKGKGIIFTRDRKPKAAFLQKRWGMN 592

RESULT 12

ADD27986 standard; protein: 604 AA.

ADD27986;

15-JAN-2004 (first entry)

Beta-glucuronidase.

beta-glucuronidase; altered seed meal; oil body; plant; enzyme.

Escherichia coli.

US2003126631-A1.

03-JUL-2003.

29-JUN-2001; 2001US-00893525.

22-FEB-1991; 91US-00659835.

16-NOV-1993; 93US-00142418.

PR 30-DEC-1994; 94US-00366783.

PR 25-APR-1997; 97US-00846021.

PR 18-DEC-1998; 98US-00210843.

(MOLO/) MOLONEY M. M.

(VROO/) VAN ROOIJEN G.

Moloney MM, Van Rooijen G;

N-PSDB; ADD27985.

Disclosure; SEQ ID NO 37; 52pp; English.

XX The invention relates to a method of expressing a heterologous
CC polypeptide by a host cell. The methods are useful for producing altered
CC seed meals by manipulating oil bodies of plants. The present sequence is
CC used in the exemplification of the present invention.
XX
SQ Sequence 604 AA:

Query Match 50.9%; Score 1706; DB 7; Length 604;

Best Local Similarity 54.8%; Pred. No. 4.6e-130;

Matches 329; Conservative 94; Mismatches 161; Indels 16; Gaps 9;

QY 40 VROPQTSRELVNLDGLMKFAL--ASGLNDTAPQWTAAPKGECPVASYNDIFISRE 96

Db 3 LRVEPTREIRKTLKLDGMAFSLRBNGIQQ--RMWESALQESALVAPSPDQFADAD 60

QY 97 IHDHVGVVYQREVIYFKGWSQERYLVRAESATHHGHYNNRLVAEHVGGYTFPEADVT 156

Db 61 IRRVAGVWVYQREVFIPKMGAGORIVLRFDVATHYGVWVNNQVMEHOGGYTFPEADVT 120

QY 157 ELVAPGKFLITGVNNELTHEITIPGKITGNATGRIQIOTYOHDFYVAGLSRWLS 216

Db 121 PYVIAGKSVRITVCNNELMWQITIPGMVIT--DENGKKKOSYFHDFPNVAGIHSRWLYT 179

QY 217 VPQOHIODITVYTDV--DCDNGLINEVEVANOTGQIQISVIDEDGAIYAKASGAGTV 274

Db 180 TPNTWDDITVYTHVADCNHASVDWQV--VAN--GDVSELDADQOVATGQTSGLT 235

QY 275 TITSVVLMOGAAVLYOLOVNISSGSDVDVTYNLATGVTVVAVSQFLINKPFFETG 334

Db 236 QVNPFLMKGEGEYLYELCV--TAKSQTECDIPLRVGIRSVAVKGOFLINHPFFETG 293

QY 335 FGHEDTAVRGKGDPAVYVHDFOLMKMGIGANSFRISHYAEVMDPADRNGIVDET 394

Db 294 FGHEDADLRGKGFVDVYVHDHADMWIGANSFRISHYAEVMDPADRNGIVDET 353

QY 395 PAVGNIAL-MGVSESGAP-QTFPPDAINDKTOEAKHQAIRELIARDKNASVVMMSIAN 452

Db 354 AAVGFSLSLIGIPEAGNKPKEYSEAVNGETOAHQAIRELIARDKNHPSVVMMSIAN 413

QY 453 EPASHEDGAREYEPFLTNLTROLDPTFPITFANVGTATYOLDRIISDLFDVSCINRYGWY 512

Db 414 EPDTRPGAREYFAPLAETARKLDPTFPITCVMMFCDAHTDITSDLDVCLNRYGWY 473

QY 513 SQTGDLSEAAALEKELHGWQEKFRPIVMTVEGADTLAHLSTGLPMSEEPQVQWLDL 572

Db 474 VQSGDLETAKEVLEKELAMOEKIHOPITTEYGVDTLAGHSVYTDMSSEETQCAWLDL 533

QY 573 YHRVFDRIESMAGBHVNFADPOTNLGIIIRVDSNKGKGFTRDRKPKAAASLRARWTSID 632

Db 534 YHRVFDRIESVAVGEQVNFADPOTNLGIIIRVDSNKGKGIIFTRDRKPKAAFLQKRWGMN 593

RESULT 13
ADD27989 standard; protein: 659 AA.

ADD27989;

15-JAN-2004 (first entry)

Oleosin/beta-glucuronidase fusion protein #2.

beta-glucuronidase; oleosin; thale cress; altered seed meal; oil body;

plant; enzyme.

Escherichia coli.

OS Arabidopsis thaliana.

US2003126631-A1.

03-JUL-2003.


```

Db      482 QVNNPRLMOPGEGYLVELCV--TAKSQTECDIYPLRVGIRSVAVKGOQFLINHKPFYFTG 539
Qy      335 FKGHEDTAVRGKGDPAVYVHDFOLMKWIGANSFRTSHYPYAEVNDPDRNGIIVIDET 394
Db      540 FGRHEDADLRGKGFDPVNLVHDLMDWIGANSYRTSHYPYAEMLMADDEHGIIVIDET 599
Qy      395 PAVGLNIAL-MGVSSSGAP-QFTTPDAINDKTOEAKHKAIRBELIARDKXHASVVMMSIAN 452
Db      600 AAVGFNLISLIGIFEAGNKKPELYSEEAUVNGETOAHQAIKELIARDKXHPSVVMSIAN 659
Qy      453 EPASHEDGAREYFEPLTNLTROLDPTPTPTFANVGTATYQLODRISDLFDVSCINRYFGWY 512
Db      660 EBDTRPQGARREYFAPLAETARKDPTPTTCVVMFCDAHTDTISDLFDVLCINRYFGWY 719
Qy      513 SQTGLEEBALEALKEKELHGMOKFHRPIVMTXGADTLAGLSILGLPMSSEFOVOMLDM 572
Db      720 VOSGDLETAKEKYLEKELHGMOKFHRPIITTEYGVDTLAGLSHMYTDMMSSEYQCAMLDM 779
Qy      573 YHRVFPDRISMAGEHYWNPADPOTNLGITRVDSNKKGVFTDRKKPKAAHSLRAFTSID 632
Db      780 YHRVFPDRISAVVGEQWNPADPOTNLGITRVDSNKKGVFTDRKKPKAAHSLRAFTSID 839

```

RESULT 15
 ABB84107
 ID ABB84107 standard; protein; 603 AA.

```

AC      ABB84107;
XX      17-SEP-2002 (first entry)
DT      GUS protein #1.
XX      GUS; refractory; beta-glucuronidase; screening; ds.
OS      Unidentified.
XX      CN1338515-A.
XX      06-MAR-2002.
XX      18-AUG-2000; 2000CN-00119633.
XX      18-AUG-2000; 2000CN-00119633.
XX      (REBI-) RES CENT BIO TECHNOLOGY SHANGHAI ACAD AG.
XX      Yao Q, Xiong A, Peng R;
XX      WPI; 2002-384278/42.
XX      N-PSDB; ABL61358.
XX      Refractory beta-glucosiduronatase gene and its obtaining process.
XX      Claim 1; Page 1-3 (Claims); 23pp; Chinese.
XX      This invention describes a novel refractory beta-glucuronidase (GUS) gene
XX      prepared through DNA mutation. The gene is used in the creation of an
XX      expression carrier, which is transferred into a colibacillus. The GUS
XX      gene has refractory power (85 degrees C) and can be used to screen
XX      transgenic plants with high efficiency. This sequence represents a GUS
XX      protein described in the disclosure of the invention
XX      SQ      Sequence 603 AA;

```

Query Match 50.6%; Score 1697; DB 5; Length 603;
 Best Local Similarity 54.5%; Pred. No.2.5e-129;
 Matches 327; Conservative 95; Mismatches 162; Indels 16; Gaps 9;

Qy 40 VNPQTSSEELVNLGDLMKFAL--ASGLNDTAQPTAPLPKGLBCPPVASYNDIFISRE 96
 Db 2 LRPVETPTREIKLQDLMAFSLDRNCGIDQ--RWMESALQESRAIAYVPSFDFADAD 59

```

Qy      97 IDHVGWVYVYOREBIVPKGMSOERYLVRAESATTHGRIVYNNRLVAEHGYGTPPEADVT 156
Db      60 IRRVAGNVYQOEVEPLPKMACGRITVLRPDATYHVKWVNNQVEVHEHGGITTFEADVT 119
Qy      157 ELVAPGEKRLTIGVNNELTHETIPPKITTGATGKRIQYQHDENVYAGLARSIMLYS 216
Db      120 PYIACKSVRITVCVNNELMWTIPFGWIT--DENCKKQSYHDFPNVAGIHRSVMLVLT 178
Qy      217 VPOQHIDITVYTDV--DDNGNLINYEVEVANOQTQIQISVIDEGALVAKASGAGTY 274
Db      179 TPVTWDDITVYTHVADCNHASVDMQV--VAN--GDVSVELADADQOVVATGQGTSGTL 234
Qy      275 TIPSVKLMOPGAAYLYOLOVNIYSSGDDVDFTYNLATGRTYKVAAGSOFPLINKPFYFTG 334
Db      235 QVNNPRLMOPGEGYLVELCV--TAKSQTECDIYPLRVGIRSVAVKGOQFLINHKPFYFTG 292
Qy      335 FKGHEDTAVRGKGDPAVYVHDFOLMKWIGANSFRTSHYPYAEVNDPDRNGIIVIDET 394
Db      293 FGRHEDADLRGKGFDPVNLVHDLMDWIGANSYRTSHYPYAEMLMADDEHGIIVIDET 352
Qy      395 PAVGLNIAL-MGVSSSGAP-QFTTPDAINDKTOEAKHKAIRBELIARDKXHASVVMMSIAN 452
Db      353 AAVGFNLISLIGIFEAGNKKPELYSEEAUVNGETOAHQAIKELIARDKXHPSVVMSIAN 412
Qy      453 EPASHEDGAREYFEPLTNLTROLDPTPTPTFANVGTATYQLODRISDLFDVSCINRYFGWY 512
Db      413 EBDTRPQGARREYFAPLAETARKDPTPTTCVVMFCDAHTDTISDLFDVLCINRYFGWY 472
Qy      513 SQTGLEEBALEALKEKELHGMOKFHRPIVMTXGADTLAGLSILGLPMSSEFOVOMLDM 572
Db      473 VOSGDLETAKEKYLEKELHGMOKFHRPIITTEYGVDTLAGLSHMYTDMMSSEYQCAMLDM 532
Qy      573 YHRVFPDRISMAGEHYWNPADPOTNLGITRVDSNKKGVFTDRKKPKAAHSLRAFTSID 632
Db      533 YHRVFPDRISAVVGEQWNPADPOTNLGITRVDSNKKGVFTDRKKPKAAHSLRAFTSID 592

```

Search completed: March 18, 2005, 23:48:06
 Job time : 174 secs

THIS PAGE BLANK (UPR)